

Prediksi Jumlah Kasus Covid-19 di Indonesia menggunakan Data Google Trends dengan Metode *Hybrid Artificial Neural Network* dan *Multiple Regression*

Ferninda Maharani Kumala¹, Indwiarti², Annisa Aditsania³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹fernindamaharani@students.telkomuniversity.ac.id, ²indwiarti@telkomuniversity.ac.id,

³aaditsania@telkomuniversity.ac.id

Abstrak

Covid-19 (*Corona Virus Disease 2019*) dikategorikan sebagai sebuah pandemi. Statistik pertumbuhan kasus covid yang semakin pesat tentu perlu adanya penanganan khusus dari Pemerintah maupun seluruh lapisan masyarakat. Salah satunya dengan melakukan langkah mitigasi yaitu dengan memprediksi kasus Positif pada kasus ini. Data yang diambil untuk melakukan prediksi terhadap kasus ini bersumber dari Google Trends karena sangat tingginya popularitas kata kunci terhadap kasus Covid-19 ini di Internet, sehingga memudahkan dalam melakukan prediksi data. Metode yang digunakan dalam penelitian ini adalah *Multiple Linear Regression* (MLR), *Artificial Neural Network* (ANN), dan *Hybrid Artificial Neural Network* dan *Multiple Linear Regression* (MRL-ANN). Dalam penelitian ini diperoleh nilai MAPE, MAE, RMSE masing masing sebesar 12.92%, 887.68, dan 1178.68 untuk metode Hybrid Artificial Neural Network dengan Multiple Linear Regression, nilai tersebut lebih kecil dibanding metode Artificial Neural Network dengan tingkat error sebesar 13.51%, 910 dan 1202.11, pada Multiple Linear Regression tingkat error sebesar 18.07%, 1342.64, dan 1707.95. Hal ini menunjukkan metode Hybrid pada penelitian di kasus ini memberikan performansi yang lebih baik.

Kata kunci: Covid-19, Prediksi, *Hybrid*, ANN, MLR, *Google Trends*

Abstract

Covid-19 (*Corona Virus Disease 2019*) is categorized as a pandemic. The rapid growth statistics of covid cases certainly require special treatment from the government and all levels of society. Taking mitigation steps is needed by forecasting positive cases in this research. The data taken to make predictions on this case comes from Google Trends because of the very high popularity of keywords for the Covid-19 case on the Internet, making it easier to predict data. The method used in this research is Multiple Linear Regression (MLR), Artificial Neural Network (ANN), and Hybrid Artificial Neural Network with Multiple Linear Regression (MRL-ANN). In this study, the values of MAPE, MAE, RMSE were obtained respectively 12.92%, 887.68, and 1178.68 for the method of Hybrid Artificial Neural Network with Multiple Linear Regression, the value is smaller than the method of Artificial Neural Network with error rate of 13.51%, 910 and 1202.11, at Multiple Linear Regression error rates of 18.07%, 1342.64, and 1707.95. This shows the Hybrid method in the method in this research gives better performance.

Keywords: Covid-19, Forecasting, Hybrid, ANN, MLR, *Google Trends*

1. Pendahuluan

Latar Belakang

World Health Organization (WHO) dalam situs resminya menyatakan bahwa COVID-19 (*Corona Virus Disease 2019*) dikategorikan sebagai sebuah pandemic [1]. Pandemi adalah epidemi (wabah) yang terjadi pada skala yang melintasi batas internasional, biasanya menjangkit sejumlah besar orang [2]. Hingga 23 Januari 2021, 10:57 pm GMT+7, tercatat ada 96.877.399 kasus positif COVID-19 di seluruh dunia, 2.098.879 diantaranya meninggal dunia [3]. Di Indonesia sendiri tercatat ada 977.474 kasus positif COVID-19 dan 27.664 diantaranya meninggal dunia [4]. Prediksi jumlah kasus COVID-19 di Indonesia untuk masa mendatang dapat menjadi suatu acuan yang dapat diterapkan dalam penanganan kasus tersebut, sehingga dapat digunakan Pemerintah, BNPB, bahkan masyarakat umum dalam mengambil suatu keputusan terkait penanganan dan antisipasi akan kasus ini. Penelitian ini, menggunakan data Google Trends dan BNPB.

Untuk menganalisis korelasi antara *search engine* dengan data aktual, kami merujuk pada penelitian sebelumnya. Google Trends adalah situs web yang dimiliki Google.Inc yang berisi trend penggunaan kata kunci di website mesin pencari google dan berita yang sedang trend. Google Trend dapat dimanfaatkan untuk menentukan topik pencarian query untuk rentang waktu dan daerah tertentu mulai dari geografis dunia hingga geografis tingkat lokal [9]. Penelitian sebelumnya menemukan korelasi sedang hingga kuat antara data yang diperoleh dari penelusuran kata kunci terkait COVID-19 di Google Trends dan total kasus COVID-19 di Amerika Serikat seperti yang diperoleh dari agregator data nasional [20].

Metode yang digunakan untuk memprediksi jumlah kasus positif COVID-19 di Indonesia kedepannya adalah dengan menggunakan Metode *Hybrid Artificial Neural Network* dan *Multiple Regression*. *Artificial Neural Network (Artificial Neural Network)* merupakan salah satu metode berbasis kecerdasan buatan (*Artificial Intelligence*), yang mampu mengidentifikasi pola, signal prosesi serta prediksi dari sistem dengan metode pembelajaran [5]. Dengan metode-metode tersebut akan dilakukan prediksi terhadap jumlah kasus positif COVID-19 di Indonesia untuk waktu kedepan dengan menggunakan data historis harian jumlah kasus positif COVID-19 yang diambil dari situs resmi BNPB dan juga Data Google Trends. Dari penelitian yang dilakukan oleh Budi (2018) diketahui bahwa Metode *Artificial Neural Network (ANN)* merupakan metode yang dapat memberikan tingkat akurasi tinggi dalam prediksi volume penggunaan air PDAM [6]. Dalam melakukan pemodelan data prediktif, metode *Multiple Linear Regression (MLR)* adalah teknik yang umum digunakan untuk mendapatkan model input output linear untuk dataset yang diberikan [7]. Namun, model ini akan menghadapi beberapa kesulitan, terutama ketika variabel independen mengikuti distribusi tertentu. Dengan demikian, digunakan pendekatan *Hybrid ANN-MLR* untuk mengekstraksi informasi karena metode ini *self-adaptif, self-organizing* dan *error-tolerance* [8]. Metode ini merupakan pendekatan dan penggabungan antara metode *Artificial Neural Network* dan *Multiple Linear Regression*, pendekatan ini dilakukan agar dapat meminimalisasi error sehingga diperoleh data dengan MAE, RMSE, dan MAPE yang semakin kecil.

Topik dan Batasannya

Penelitian tugas akhir ini melakukan hibridasi pada metode *Multiple Linear Regression (MLR)* dan *Artificial Neural Network (ANN)* untuk melakukan *forecasting* dengan hasil yang lebih baik dibanding MLR dan ANN, yaitu *forecasting* dengan nilai error yang lebih kecil. Data yang digunakan berasal dari data harian Covid-19 di Indonesia yang bersumber dari BNPB dan data *Google Trends* harian. Penelitian ini mengukur tingkat error pada *Multiple Linear Regression, Artificial Neural Network* serta *Hybrid Multiple Linear Regression dengan Artificial Neural Network (MLR-ANN)*.

Tujuan

Tujuan penelitian tugas akhir ini adalah untuk mengetahui model atau metode *forecasting* yang paling baik untuk memprediksi kasus positif Covid-19 di Indonesia.

Organisasi Tulisan

Urutan penulisan pada tugas akhir ini adalah sebagai berikut: bagian 2 menjelaskan penelitian-penelitian terkait metode MLR, ANN, dan MLR-ANN yang telah dilakukan sebelumnya. Bagian 3 menjelaskan mengenai metode yang dibangun serta alur perancangannya. Bagian 4 menjelaskan metode yang dilakukan dan hasil yang diperoleh dari tiap metode. Bagian 5 menjelaskan kesimpulan dari penelitian tugas akhir ini dan saran yang dapat digunakan untuk penelitian selanjutnya.

2. Studi Terkait

Google Trends

Google Trends adalah situs web yang dimiliki Google.Inc yang berisi trend penggunaan kata kunci di website mesin pencari google dan berita yang sedang trend. Google Trend dapat dimanfaatkan untuk menentukan topik pencarian query untuk rentang waktu dan daerah tertentu mulai dari geografis dunia hingga geografis tingkat lokal. [9]. Perkembangan era ke arah digital berakibat terdapatnya berbagai jenis data yang beragam mulai dari teks, foto, video, musik, postingan di sosial media dan memiliki karakteristik 4V (*Volume, Velocity, Variety, Veracity*) yang dikenal dengan nama Big Data. Peningkatan penggunaan data Google Trends ditemukan dalam berbagai domain, sehingga Google Trends tidak hanya membantu peneliti untuk menyelidiki berbagai masalah masa lalu dan sekarang yang terkait dengan domain yang tercantum di atas, tetapi juga memprediksi suatu hal di masa depan [10]. Data tersebut dapat digunakan sebagai data penelitian yang dapat meningkatkan tingkat keakuratan dalam menyelesaikan suatu permasalahan. Sebagai salah satu big data sederhana, Google Trends dapat digunakan sebagai prediksi maupun kebijakan lainnya dan pada penelitian lain disebutkan bahwa korelasi *query* pada Google Trends memberikan hasil yang baik dengan data aktual [20].

Metode *Artificial Neural Network (ANN)*

Pada penelitian [11] ANN telah berhasil diterapkan dalam peramalan *time series* dalam bidang pengetahuan yang berbeda seperti biologi, keuangan dan ekonomi, konsumsi energi, obat-obatan, meteorologi dan pariwisata. *Backpropagation* merupakan algoritma pembelajaran yang terawasi dan biasanya digunakan oleh perceptron dengan banyak lapisan untuk mengubah bobot-bobot yang terhubung dengan neuron-neuron yang ada pada *hidden layer*-nya [11]. *Backpropagation* model yang paling banyak digunakan untuk pemodelan, prediksi, dan klasifikasi. Model ini ditandai dengan serangkaian unit pemrosesan tiga lapis yang terhubung dengan tautan asiklik. Algoritma

pembelajaran untuk ANN *Backpropagation* adalah sebagai berikut [12]. Hubungan antara output (y) dan input ($\phi_1, \phi_2, \dots, \phi_p$) yang dapat diklasifikasikan ke dalam Persamaan matematika ini (1) [13] adalah sebagai berikut:

(1) Dimana ϕ_i adalah bobot nilai inputnya, $i=1, 2, \dots, p$, $j=1, 2, \dots, q$ adalah model parameter yang dikenal bobot koneksi, p dan q adalah jumlah input dan jumlah node tersembunyi, g adalah fungsi transfer / aktivasi, dan ϵ adalah input error.

Data yang dimasukkan akan melewati lapisan input dari jaringan syaraf. Kemudian akan melewati lapisan tersembunyi dan keluar melalui lapisan output. Setiap simpul lapisan tersembunyi dan lapisan keluaran akan mengumpulkan data dari setiap simpul di sana (baik lapisan input atau lapisan tersembunyi) dan digunakan sebagai fungsi aktivasi seperti dalam Persamaan (2) [13]:

$$g(x) = \max\{0, x\} \tag{2}$$

Salah satu jenis fungsi aktivasi ditunjukkan oleh neuron dalam jaringan untuk model ANN ditunjukkan dalam Persamaan (3) yaitu:

$$Y = \phi(\phi_{i-1}, \dots, \phi_{H-1}, \phi_i) + \epsilon \tag{3}$$

di mana, w adalah vektor untuk semua parameter, ϕ_i adalah variabel prediktif masa mendatang, $f(\cdot)$ dan ϵ adalah fungsi untuk struktur jaringan dan bobot dan kesalahan koneksi.

Metode Multiple Linear Regression (MLR)

Regresi Linier Berganda adalah hubungan antara variabel dependen y dan banyak variabel independen ($\phi_1, \phi_2, \dots, \phi_p$). Regresi linier hanya memiliki satu variabel dependen yang bervariasi dengan satu variabel independen. Namun, ketika kita memiliki variabel dependen dengan dua atau lebih variabel independen, maka kita perlu menggunakan regresi linier berganda. Tujuan dari MLR adalah untuk menjelaskan sebanyak mungkin variasi yang diamati dalam *Time Series* kasus positif COVID-19 [14]. Untuk data *time series*, model regresi linier berganda dapat dinotasikan seperti ini:

$$Y = \phi_0 + \phi_1 \phi_1(t) + \phi_2 \phi_2(t) + \phi_3 \phi_3(t) + \dots + \phi_n \phi_n(t) + \epsilon \tag{4}$$

di mana, ϕ_i adalah koefisien parameter, $y(t)$ adalah variabel dependen atau variabel yang akan diprediksi dan ϕ_i variabel independen atau variabel prediktor dan ϵ adalah error pada waktu t.

Metode Hybrid ANN-MLR

Pada metode Hybrid *Multiple Linear Regression* dengan *Artificial Neural Network* ini, pertama dilakukan prediksi terhadap variabel ϕ_i terlebih dahulu dengan metode *Multiple Linear Regression* dengan menggunakan persamaan (4), setelah itu hasil residual atau *gap error* dari MLR digunakan kembali saat proses training dengan *Artificial Neural Network*. Karena model MLR tidak dapat mendeteksi struktur non-linier dari variabel untuk prediksi kasus Covid-19 padahal residual model linier berisi informasi tentang non-linieritas [23]. Agar menghasilkan model yang lebih umum, model hybrid linear dan model non-linear yang lebih akurat, dibuat persamaan untuk model hybrid ANN-MLR. Di model ini *time series* juga dianggap sebagai fungsi linear dan non-linear. Fungsi persamaan sebagai berikut:

$$Y = l_t + \phi \tag{5}$$

di mana l_t adalah komponen linear dan ϕ adalah komponen non-linear. Di tahap pertama, tujuan utamanya adalah untuk mendapatkan komponen linear yang merupakan model MLR. Kemudian pada tahap ini akan dibuat persamaan ϵ_t yang merupakan error dari waktu t pada model linear:

$$\epsilon_t = Y - \hat{Y} \tag{6}$$

\hat{Y} adalah Variabel perkiraan pada waktu t. Hasil dari prediksi dan error pada linear modelling adalah hasil dari tahap pertama dan akan digunakan untuk tahap selanjutnya yaitu dengan metode ANN. Pada tahap kedua, fokus utama adalah pada model non linier. *Multi-layer perceptron* digunakan untuk memodelkan hubungan non-linear. Dengan n sebagai input node, model untuk error ANN adalah:

$$\epsilon_t = \phi(\phi_{H-1}, \phi_{H-2}, \dots, \phi_{H-1}) + \epsilon \tag{7}$$

dimana $\phi(\cdot)$ adalah fungsi non-linear untuk menentukan neural network and ϵ adalah nilai error random.

Perbandingan Kinerja Metode

Untuk membandingkan kinerja metode umumnya divalidasi dengan menggunakan sejumlah indikator. Pada penelitian ini digunakan indikator *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE). Menurut estimasi error, semakin kecil error, semakin tinggi akurasi data. Model evaluasi kriteria kinerja untuk mengukur error adalah sebagai berikut:

$$RMSE = \sqrt{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 / n} \tag{9}$$

$$MAE = \sum_{i=1}^n |Y_i - \hat{Y}_i| / n \tag{10}$$

$$MAPE = \frac{\sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}}{n} \times 100\% \quad (11)$$

3. Data dan Metodologi

Data

Penelitian ini menggunakan data historis Covid-19 di Indonesia yang diunduh melalui website BNPB mulai tanggal 2 Maret 2020 yang ditandai dengan kasus pertama positif di Indonesia hingga tanggal 16 Januari 2021. Data diunduh pada tanggal 17 Januari 2021 pada laman <https://bnpb-inacovid19.hub.arcgis.com> yang tersedia secara *online*. Di penelitian ini ditambahkan beberapa *query* Google Trends yang merujuk kepada Covid-19 diantaranya “PSBB” dan “Virus Corona”. Data google trends diunduh melalui <https://trends.google.com>. Pengambilan kata kunci *Google Trends* merujuk pada *top searches* google 2020 [21]. Data yang diunduh pada google trends berlokasi di Indonesia dan diunduh untuk rentang waktu 2 Maret 2020 hingga 16 Januari 2021. Data dari kedua sumber tersebut kemudian dibagi menjadi 2 yaitu data *training* dan data *testing*. Data *training* diambil dari tanggal 2 Maret 2020 hingga tanggal 13 November 2020 dan data *Testing* diambil dari tanggal 14 November 2020 hingga tanggal 16 Januari 2021. Pada penelitian Data *training* digunakan untuk melatih model, dan data *testing* digunakan untuk mengukur performansi model. Dengan rasio pembagian dataset sebesar 80:20 untuk data *training* dan *testing*. Pada penelitian ini akan di prediksi variabel y “Positif” dengan variabel x “Meninggal”, “Sembuh”, “Virus Corona”, dan “PSBB” sebagai prediktor. Pembagian dataset untuk penelitian ini dapat dilihat pada Tabel 1.

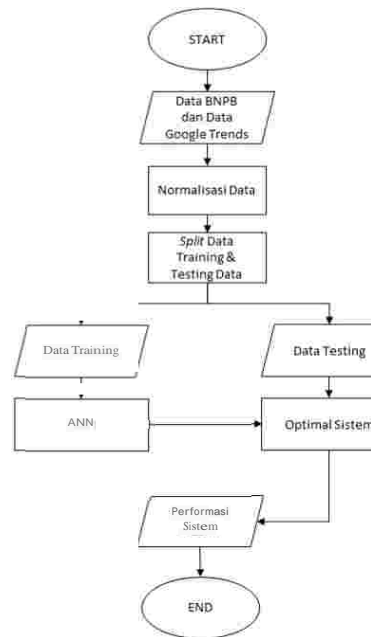
| Variabel | Nama Variabel | Keterangan |
|----------|---------------|--|
| Y | Positif | Jumlah data orang terkonfirmasi “Positif” covid harian |
| X1 | Sembuh | Jumlah data orang terkonfirmasi “Sembuh” covid harian |
| X2 | Meninggal | Jumlah data orang terkonfirmasi “Meninggal” covid harian |
| X3 | Virus Corona | Jumlah <i>trends keyword</i> “Virus Corona” harian |
| X4 | PSBB | Jumlah <i>trends keyword</i> “PSBB” harian |

Tabel 1. Dimensi Dataset yang digunakan

Artificial Neural Network (ANN)

Dalam prediksi kasus COVID-19 di Indonesia ini digunakan algoritma *backpropagation* dan juga menggunakan arsitektur jaringan *multilayer network*. Kelebihan arsitektur jaringan ini adalah dapat menyelesaikan masalah yang lebih sulit dibandingkan dengan *single-layer* [17]. Pada metode ini normalisasi data sering dilakukan sebelum proses pelatihan dimulai untuk menstandarisasi output dan input dan menghindari masalah komputasi. Untuk membatasi atau menekan output yang mungkin dari node, biasanya node dibuat (0,1) atau (-1,1) [19]. Setelah itu dieksekusi dengan algoritma *backpropagation* ada beberapa tahapan yaitu:

1. Melatih data *Training* untuk dilatih polanya, melakukan inisiasi bobot secara acak dengan bilangan acak terkecil.
2. Kemudian sinyal dipropagasikan ke hidden layer menggunakan fungsi aktivasi tertentu, output jaringan data train (dt) dibandingkan dengan target data testing yang harus dicapai (dv), setelah itu dihitung selisih dt-dv.
3. Kemudian y didistribusikan ke semua unit yang terhubung. Dalam algoritma ini dipilih data *training* yang sudah dipisahkan pada bagian *processing* data yang akan dilatih untuk menjadi output jaringan data train (dt) kemudian dibandingkan dengan data *testing* untuk menjadi data target.
4. Setelah itu hasil training data dibandingkan dengan data target yaitu data *testing* untuk mendapatkan analisis performansi ANN yaitu RMSE, MAPE, dan MAE.
5. Lalu dilihat hasil training data terhadap *forecasting data* untuk mendapatkan prediksi terhadap kasus positif Covid.

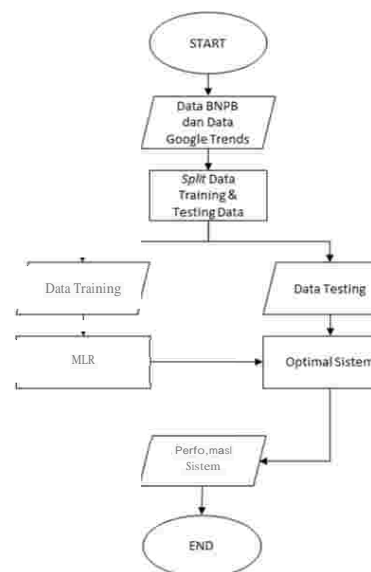


Gambar 1. flowchart Alur Penelitian dengan metode ANN

Multiple Linear Regression (MLR)

Pada penelitian ini untuk mendapatkan hasil prediksi dengan metode *Multiple Linear Regression*, variabel y dependen adalah variabel kasus “Positif”, dan untuk variabel x independen adalah kasus “Sembuh”, “Meninggal”, “Virus Corona”, dan “PSBB”. Berikut langkah-langkah yang dilakukan pada metode ini:

1. Baris data yang menjadi data *training* dieksekusi dengan model MLR pertama dengan membuat data *constant* setiap variabel data x
2. Setelah itu data dilatih menggunakan model MLR lalu kemudian akan terlihat koefisien untuk tiap variabel independen x terhadap variabel dependen y
3. Data yang sudah di training dengan MLR di validasi dengan *testing* data untuk mendapatkan hasil performansi model yaitu RMSE, MAE, dan MAPE.
4. Kemudian setelah itu, data yang sudah di training digunakan untuk melakukan forecasting.

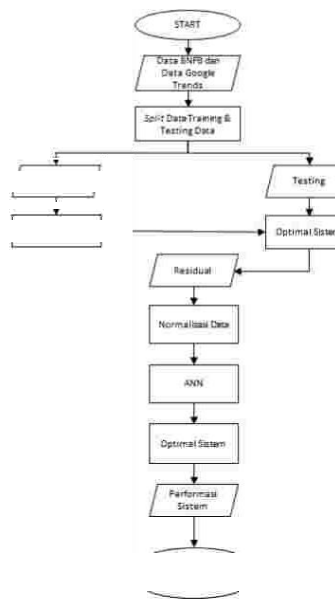


Gambar 2. flowchart Alur Penelitian dengan metode MLR

Hybrid Artificial Neural Network dan Multiple Linear Regression (ANN-MLR)

Pada proses hybrid antara ANN dan MLR yang pertama dilakukan adalah melakukan pelatihan data untuk mendapatkan hasil prediksi dengan metode *Multiple Linear Regression*, data y dependen adalah variabel kasus “Positif”, dan untuk variabel independen x adalah data “Sembuh”, “Meninggal”, “Positif Covid”, dan “Covid”. Berikut tahapan yang dilakukan:

1. Baris data yang menjadi data training dieksekusi dengan model MLR, dengan membuat data *constant* setiap variabel data x kemudian setelah itu data di latih menggunakan model MLR lalu kemudian akan terlihat koefisien untuk tiap variabel independen x terhadap variabel dependen y
2. Kemudian data yang sudah dilatih dengan MLR divalidasi dengan data *testing* untuk mendapatkan data hasil prediksi kemudian (data y_{new}) data hasil prediksi atau validasi tersebut dinormalisasi.
3. Setelah itu di eksekusi dengan algoritma *backpropagation* melalui beberapa tahapan yaitu melatih data *training* untuk dilatih polanya, melakukan inisiasi bobot secara acak dengan bilangan acak terkecil,
4. Kemudian sinyal dipropagasikan ke hidden layer menggunakan fungsi aktivasi. Di tahap ini output jaringan yang digunakan adalah data *train* dari data y_{new} ($dt_{y_{new}}$) dibandingkan dengan target data *testing* yang juga diambil dari data y_{new} yang harus dicapai ($dv_{y_{new}}$), setelah itu dihitung selisih $dt_{y_{new}} - dv_{y_{new}}$, kemudian $dv_{y_{new}}$ didistribusikan ke semua unit yang terhubung.
5. Dalam algoritma ini dipilih data *training* yang sudah dipisahkan pada bagian *processing* data yang akan dilatih untuk menjadi output jaringan data train y_{new} ($dt_{y_{new}}$) kemudian dibandingkan dengan data *validation* y_{new} (y_{new}) untuk menjadi data target
6. Setelah itu hasil training data dibandingkan dengan data target yaitu data *testing* untuk mendapatkan analisis performansi ANN yaitu RMSE, MAPE, dan MAE. Setelah itu dilihat hasil training data terhadap *forecasting data* untuk mendapatkan prediksi terhadap kasus positif Covid.

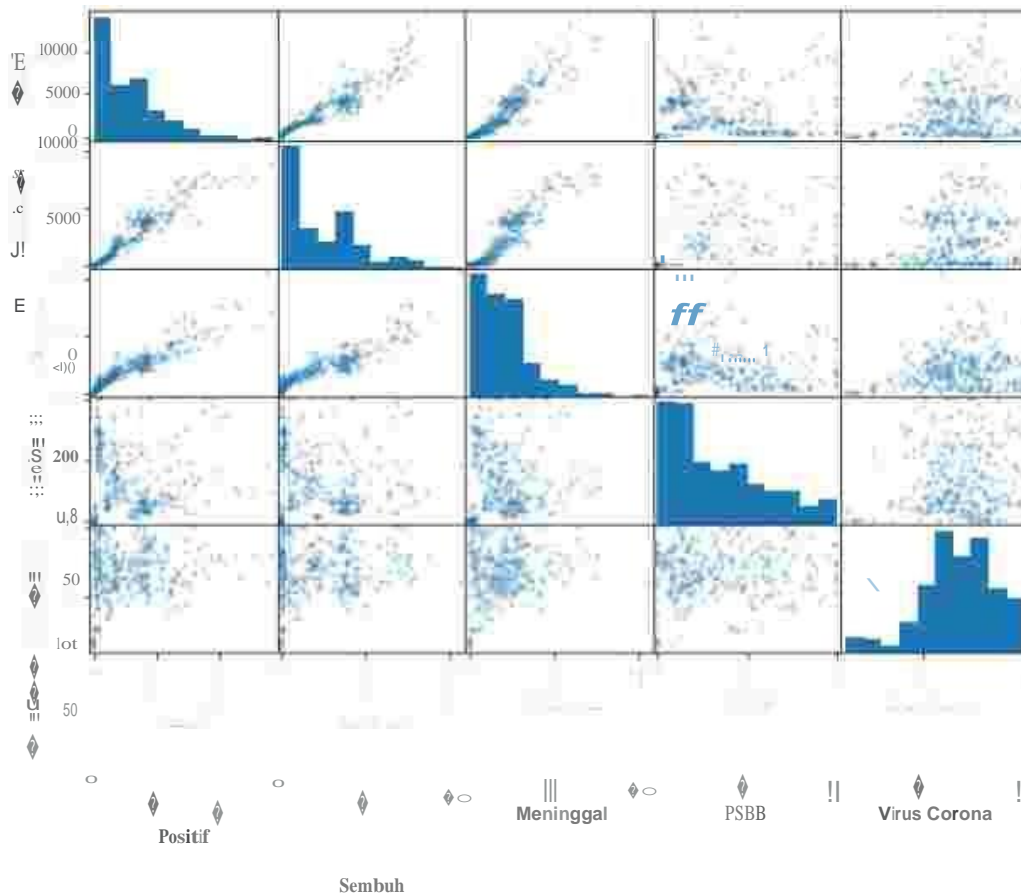


Gambar 3. flowchart Alur Penelitian dengan metode Hybrid ANN-MLR

4. Hasil dan Diskusi

Data

Pada penelitian ini digunakan metode *Multiple Linear Regression*, maka perlu dilihat hubungan linearitas antar variabel. Pada Gambar 4. terlihat yang berhubungan linear dengan variabel “Positif” adalah variabel “Sembuh” dan “Meninggal”, sekilas pada variabel google trends tidak ada hubungan linearitasnya, maka kita perlu mengetahui korelasi antar variabel, seeperti ditunjukkan pada Tabel 2.



Gambar 4. Plot hubungan linear antar variabel

| | Positif | Meninggal | Sembuh | Virus Corona | PSBB |
|--------------|---------|-----------|--------|--------------|--------|
| Positif | 1 | 0,95 | 0,94 | 0,15 | -0,031 |
| Meninggal | 0,95 | 1 | 0,92 | 0,15 | -0,088 |
| Sembuh | 0,94 | 0,92 | 1 | 0,17 | -0,027 |
| Virus Corona | 0,15 | 0,15 | 0,17 | 1 | 0,085 |
| PSBB | -0,031 | -0,088 | -0,027 | 0,085 | 1 |

Tabel 2. Korelasi Antar Variabel

Karena variabel “PSBB” berkorelasi negatif atau sangat mendekati 0, maka variabel PSBB tidak digunakan dalam prediksi variabel “Positif”.

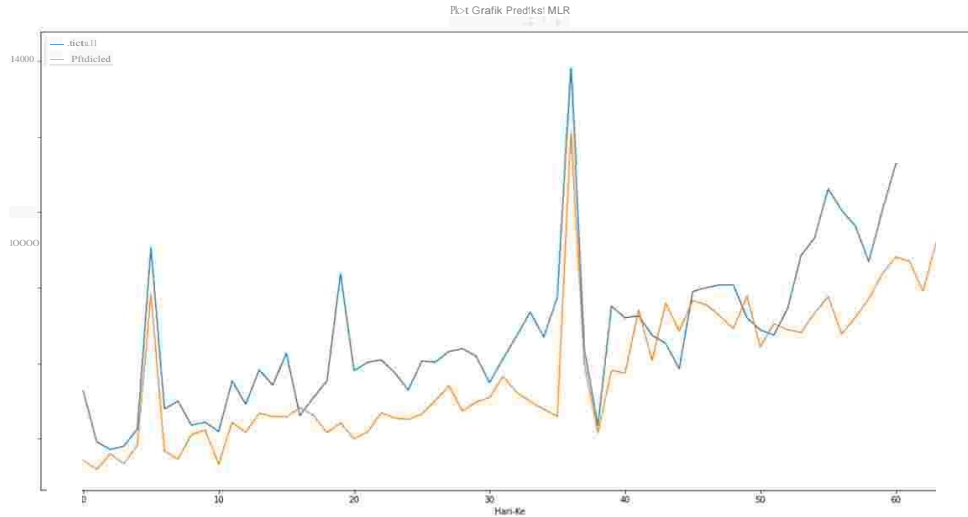
Multiple Linear Regression (MLR)

Pada tahap pengujian dengan MLR, pertama dilakukan pemisahan data Y dan X, disini yang menjadi tujuan prediksi pada penelitian ini adalah kasus “Positif” Covid-19 di Indonesia (Y) merupakan variabel dependen dan yang menjadi variabel independen atau prediktor adalah variabel “Sembuh”, “Meninggal”, “Virus Corona” yang merupakan X1, X2, dan X3. Nilai dari t_{hitung} yang lebih besar dan dua kali dari kesalahan standar (*standar error*), data tersebut dianggap sebagai data yang memiliki pengaruh signifikan [22]. *Standard error* merupakan penyimpangan dari konstanta yang ada dalam model persamaan regresi, semakin kecil penyimpangan dalam koefisien regresi, berarti semakin besar kontribusi variabel X. *Standard error* Pada penelitian ini dari 3 variabel independen, yaitu Sembuh, Meninggal, dan Virus Corona dilihat melalui koefisien regresi lebih besar dari dua kali *standard error*-nya seperti yang terlihat pada Tabel 2. Hasil ini menunjukkan bahwa 3 dari 4 data yaitu variabel “Sembuh”, “Meninggal”, dan “Virus Corona” merupakan variabel penting untuk memprediksi kasus positif Covid-19. Dari hasil korelasi koefisien data ini juga terlihat bahwa salah satu variabel memiliki koefisien cukup besar setelah variabel “Meninggal”.

| Constant | Variabel | | | |
|----------------|-----------|--------|-----------|--------------|
| | Const. | Sembuh | Meninggal | Virus Corona |
| t_{hitung} | -100.1206 | 0.6682 | 11.6794 | 2.9269 |
| Standard Error | | 0.037 | 1.428 | 1.507 |

Tabel 3. Koefisien dan Standar Error 4 Variabel Independen dengan MLR

Gambar 5 merupakan plot data antara data prediksi dengan data aktual setelah dilatih dengan model *Multiple Linear Regression*. Tabel 4. Merupakan selisih antara data aktual dengan data hasil prediksi pada MLR.



Gambar 5. Grafik Data Prediksi dan Aktual kasus Covid-19 dengan MLR

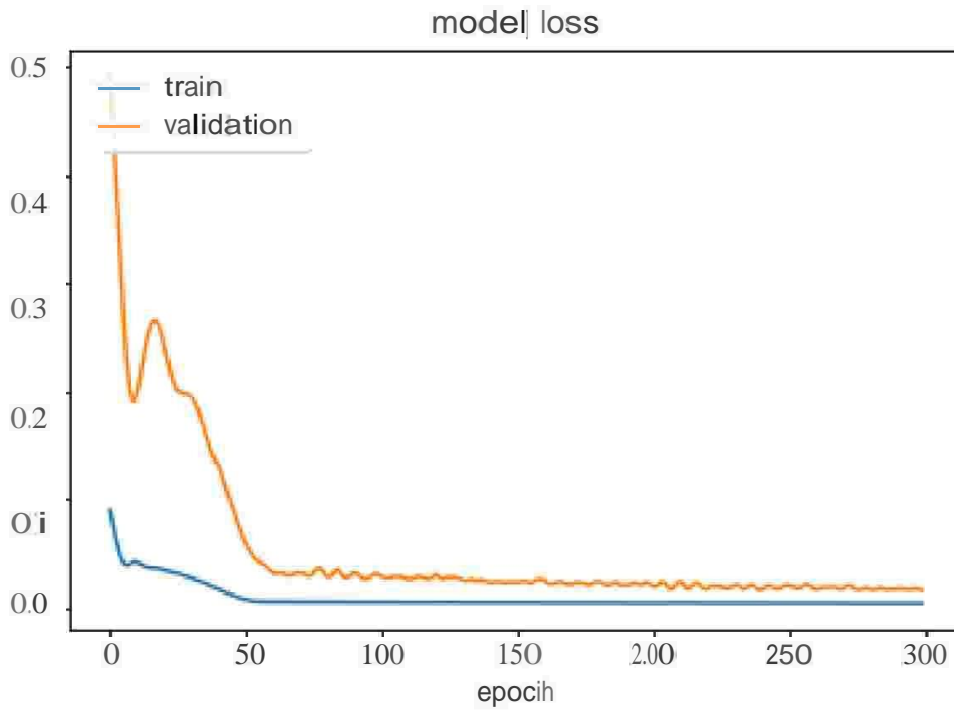
| No. | Aktual | Predicted | Aktual-Predicted |
|-----|--------|------------|------------------|
| 1 | 5272 | 3420,37713 | 1851,622866 |
| 2 | 3923 | 3196,03875 | 726,9612522 |
| 3 | 3718 | 3605,72358 | 112,2764165 |
| 4 | 3807 | 3341,92158 | 465,0784151 |
| 5 | 4265 | 3831,09441 | 433,9055936 |
| 6 | 9063 | 7834,3065 | 1228,693497 |
| 7 | 4792 | 3686,46812 | 1105,531883 |
| 8 | 4998 | 3458,85385 | 1539,146146 |
| 9 | 4360 | 4165,2537 | 194,7463038 |
| 10 | 4442 | 4249,93711 | 192,0628931 |
| 11 | 4192 | 3321,89826 | 870,1017422 |
| 12 | 5534 | 4439,05302 | 1094,946978 |
| 13 | 4917 | 4166,93484 | 750,0651558 |
| 14 | 5828 | 4669,20694 | 1158,793059 |
| 15 | 5418 | 4589,5768 | 828,4231952 |
| 16 | 6267 | 4571,69597 | 1695,304025 |
| 17 | 4617 | 4832,95973 | 215,959729 |
| 18 | 5092 | 4633,4736 | 458,5263979 |
| 19 | 5533 | 4168,06191 | 1364,938085 |
| 20 | 8369 | 4421,98352 | 3947,016479 |
| 21 | 5803 | 4019,09608 | 1783,903916 |
| 22 | 6027 | 4193,57175 | 1833,428248 |
| 23 | 6089 | 4729,92086 | 1359,079144 |
| 24 | 5754 | 4566,35119 | 1187,648806 |
| 25 | 5292 | 4513,35799 | 778,6420147 |
| 26 | 6058 | 4652,14157 | 1405,85843 |
| 27 | 6033 | 5003,14645 | 1029,853552 |
| 28 | 6310 | 5409,64323 | 900,3567736 |
| 29 | 6388 | 4734,13469 | 1653,865312 |
| 30 | 6189 | 4970,97966 | 1218,020339 |

| | | | |
|----|-------|------------|-------------|
| 31 | 5489 | 5100,2909 | 388,7090977 |
| 32 | 6120 | 5679,17129 | 440,8287148 |
| 33 | 6725 | 5250,31344 | 1474,686555 |
| 34 | 7354 | 5024,73859 | 2329,261412 |
| 35 | 6689 | 4822,68716 | 1866,312843 |
| 36 | 7751 | 4604,19609 | 3146,803906 |
| 37 | 13830 | 12146,8251 | 1683,174939 |
| 38 | 6347 | 5988,16102 | 358,838977 |
| 39 | 4360 | 4165,2537 | 194,7463038 |
| 40 | 7514 | 5835,51715 | 1678,482845 |
| 41 | 7199 | 5747,69181 | 1451,308189 |
| 42 | 7259 | 7405,13792 | 146,1379214 |
| 43 | 6740 | 6081,90852 | 658,0914837 |
| 44 | 6528 | 7605,88988 | 1077,889885 |
| 45 | 5854 | 6858,95357 | 1004,953568 |
| 46 | 7903 | 7671,12258 | 231,8774201 |
| 47 | 8002 | 7559,97251 | 442,0274949 |
| 48 | 8074 | 7274,05162 | 799,9483806 |
| 49 | 8072 | 6916,9758 | 1155,024203 |
| 50 | 7203 | 7778,31489 | 575,3148853 |
| 51 | 6877 | 6452,28009 | 424,7199144 |
| 52 | 6753 | 7048,06141 | 295,0614077 |
| 53 | 7445 | 6891,18237 | 553,8176309 |
| 54 | 8854 | 6860,19868 | 1993,801324 |
| 55 | 9321 | 7370,90023 | 1950,09977 |
| 56 | 10617 | 7810,17428 | 2806,825723 |
| 57 | 10046 | 6816,88032 | 3229,119679 |
| 58 | 9640 | 7265,14641 | 2374,853592 |
| 59 | 8692 | 7756,3001 | 935,6999007 |
| 60 | 10047 | 8381,04074 | 1665,959263 |
| 61 | 11278 | 8818,39397 | 2459,606025 |
| 62 | 11557 | 8702,14423 | 2854,855769 |
| 63 | 12818 | 7919,1283 | 4898,871702 |
| 64 | 14224 | 9227,14976 | 4996,850236 |

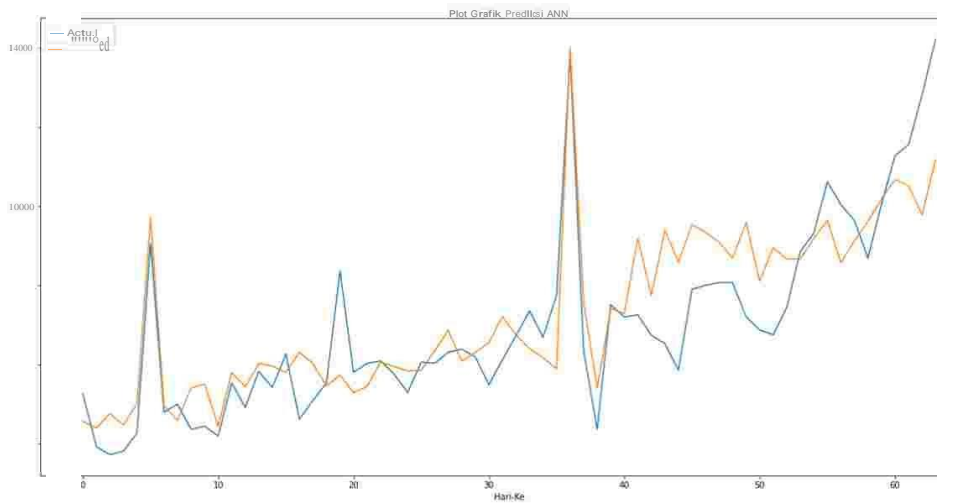
Tabel 4. Nilai selisih data *Actual* dan *Predicted* pada MLR

Artificial Neural Network (ANN)

Pada metode artificial neural network, data yang telah dibagi menjadi data *training* dan *testing* dinormalisasi atau dilakukan *scaling* data terlebih dahulu pada data *y* dinormalisasi dengan node (-1,1), kemudian input neuron diinisiasi dengan 3 sesuai dengan data prediktor *x* yang ada yaitu “Sembuh”, “Meninggal”, dan “Virus Corona” lalu output neuron diinisiasi menjadi 1 sesuai data *y* yang akan diprediksi. Setelah semua telah diinisiasi, pelatihan dengan *Artificial Neural Network* dimulai dengan inisiasi epoch sebesar 300, epoch berpengaruh saat melakukan iterasi semakin besar nilai epoch, model semakin “belajar” lebih sering yang akan menentukan akurasi dari ANN ini sendiri. Pada Gambar 6 terlihat grafik antara loss dan nilai epoch setiap iterasi. Dengan hasil training model tersebut, terlihat loss validation terhadap train validation. Pada plot ini terlihat *training loss* kurang dari *validation loss* yang menunjukkan model ini sangat cocok dengan data latih tetapi belum optimal dengan *loss validation*, dengan kata lain model tidak menggeneralisasi dengan benar ke data yang tidak terlihat atau belum optimal. seperti pada Gambar 7. Pada Tabel terlihat selisih antara data aktual dengan data hasil prediksi pada ANN.



Gambar 6. Grafik antara loss dan nilai epoch ANN



Gambar 7. Grafik Data Prediksi dan Aktual kasus Covid-19 dengan MLR

| No | Actual | Predicted | Actual-Predicted |
|----|--------|-----------|------------------|
| 1 | 5272 | 4565,903 | 706,097 |
| 2 | 3923 | 4389,337 | 466,337 |
| 3 | 3718 | 4762,3267 | 1044,3267 |
| 4 | 3807 | 4466,825 | 659,825 |
| 5 | 4265 | 4999,668 | 734,668 |
| 6 | 9063 | 9707,559 | 644,559 |
| 7 | 4792 | 4958,0513 | 166,0513 |
| 8 | 4998 | 4585,341 | 412,659 |
| 9 | 4360 | 5404,9146 | 1044,9146 |

| | | | |
|----|-------|-----------|-----------|
| 10 | 4442 | 5510,0522 | 1068,0522 |
| 11 | 4192 | 4439,6284 | 247,6284 |
| 12 | 5534 | 5793,837 | 259,837 |
| 13 | 4917 | 5437,6934 | 520,6934 |
| 14 | 5828 | 6020,332 | 192,332 |
| 15 | 5418 | 5961,0547 | 543,0547 |
| 16 | 6267 | 5798,005 | 468,995 |
| 17 | 4617 | 6309,7607 | 1692,7607 |
| 18 | 5092 | 6020,696 | 928,696 |
| 19 | 5533 | 5457,0586 | 75,9414 |
| 20 | 8369 | 5729,715 | 2639,285 |
| 21 | 5803 | 5290,1797 | 512,8203 |
| 22 | 6027 | 5443,4355 | 583,5645 |
| 23 | 6089 | 6059,1514 | 29,8486 |
| 24 | 5754 | 5945,4897 | 191,4897 |
| 25 | 5292 | 5834,6455 | 542,6455 |
| 26 | 6058 | 5859,8887 | 198,1113 |
| 27 | 6033 | 6345,2305 | 312,2305 |
| 28 | 6310 | 6863,5986 | 553,5986 |
| 29 | 6388 | 6079,4604 | 308,5396 |
| 30 | 6189 | 6305,748 | 116,748 |
| 31 | 5489 | 6552,497 | 1063,497 |
| 32 | 6120 | 7206,3555 | 1086,3555 |
| 33 | 6725 | 6749,1553 | 24,1553 |
| 34 | 7354 | 6397,6885 | 956,3115 |
| 35 | 6689 | 6173,8076 | 515,1924 |
| 36 | 7751 | 5895,515 | 1855,485 |
| 37 | 13830 | 14028,808 | 198,808 |
| 38 | 6347 | 7565,413 | 1218,413 |
| 39 | 4360 | 5404,9146 | 1044,9146 |
| 40 | 7514 | 7417,8574 | 96,1426 |
| 41 | 7199 | 7284,744 | 85,744 |
| 42 | 7259 | 9197,424 | 1938,424 |
| 43 | 6740 | 7744,2446 | 1004,2446 |
| 44 | 6528 | 9398,5205 | 2870,5205 |
| 45 | 5854 | 8595,515 | 2741,515 |
| 46 | 7903 | 9539,53 | 1636,53 |
| 47 | 8002 | 9344,659 | 1342,659 |
| 48 | 8074 | 9093,011 | 1019,011 |
| 49 | 8072 | 8692,261 | 620,261 |
| 50 | 7203 | 9597,192 | 2394,192 |
| 51 | 6877 | 8109,22 | 1232,22 |
| 52 | 6753 | 8950,342 | 2197,342 |
| 53 | 7445 | 8670,164 | 1225,164 |
| 54 | 8854 | 8667,921 | 186,079 |
| 55 | 9321 | 9184,834 | 136,166 |

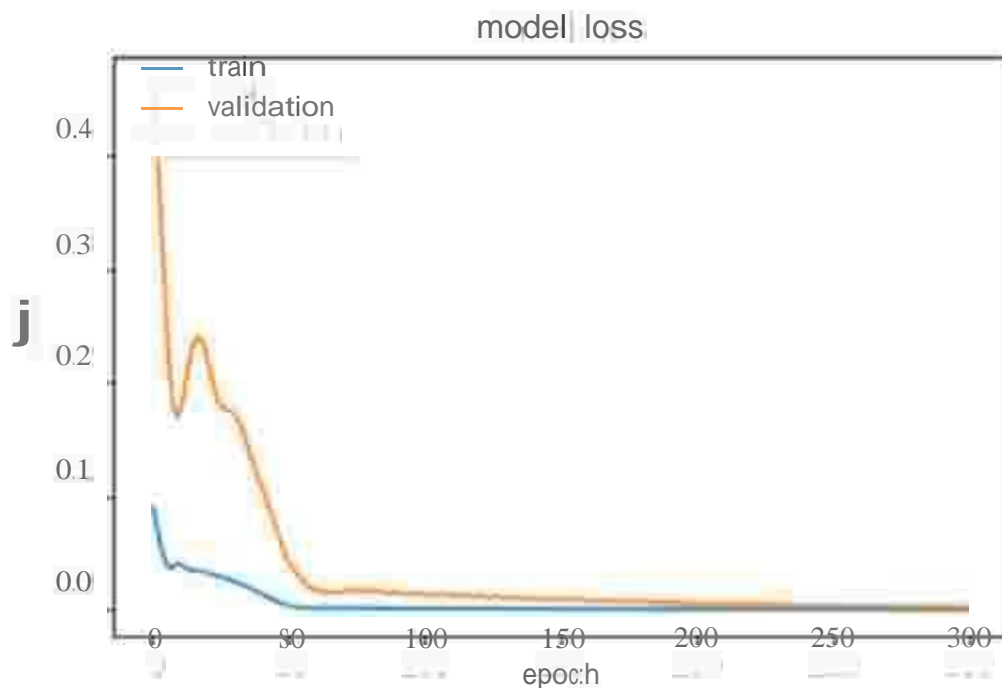
| | | | |
|----|-------|-----------|----------|
| 56 | 10617 | 9657,495 | 959,505 |
| 57 | 10046 | 8570,742 | 1475,258 |
| 58 | 9640 | 9130,098 | 509,902 |
| 59 | 8692 | 9620,191 | 928,191 |
| 60 | 10047 | 10193,827 | 146,827 |
| 61 | 11278 | 10677,463 | 600,537 |
| 62 | 11557 | 10526,477 | 1030,523 |
| 63 | 12818 | 9789,913 | 3028,087 |
| 64 | 14224 | 11178,89 | 3045,11 |

Tabel 5. Nilai selisih data *Actual* dan *Predicted* pada ANN

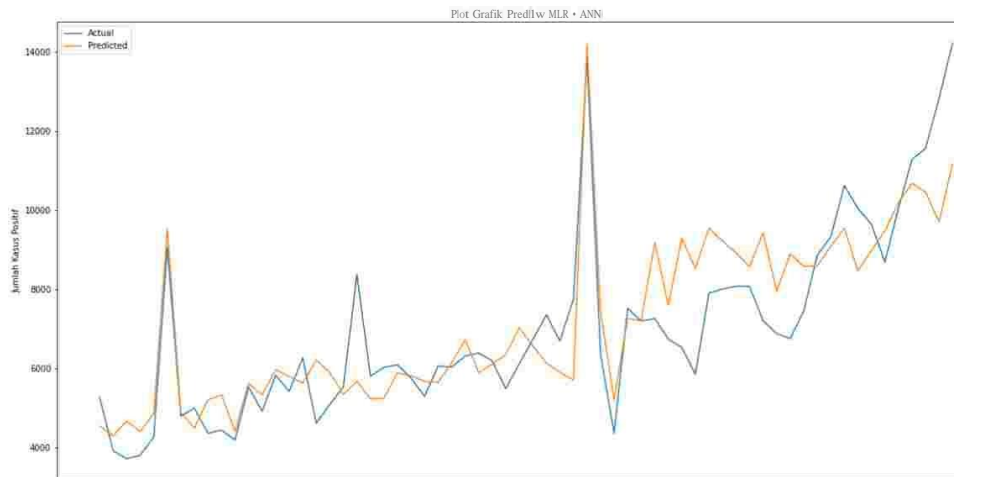
Hybrid Artificial Neural Network dan Multiple Linear Regression (ANN-MLR)

Metode hybrid pada penelitian ini dilakukan dengan mengeksekusi model linearnya dengan MLR yaitu dengan melakukan prediksi y dengan MLR, setelah itu didapat hasil y baru dan juga perhitungan residual antara y data *testing* dan y data *training*. Pada Gambar 9 terlihat plot residual dari data Train yang dihasilkan dengan metode MLR dan residual data *testing* pada Gambar 5.

Untuk hasil iterasi epoch dengan nilai yang sama dengan model MLR juga terlihat pada Gambar 8. bahwa hasil iterasi epoch dengan metode hybrid lebih baik daripada dengan model ANN, terlihat bahwa semakin lama iterasi dilakukan, *validation loss* terhadap *training loss* saat training semakin dekat ke angka 0. Pada plot antara *training loss* dan *validation loss* kedua nilai akhirnya kurang lebih sama dan juga jika nilainya bertemu (plot loss dari iterasi ke iterasi) maka hal ini menunjukkan model yang optimal. Dan untuk plot antara data aktual dan prediksi dapat terlihat pada Gambar 9, serta nilai selisih data aktual dan prediksi dapat dilihat pada Tabel 6.



Gambar 8. Grafik antara loss dan nilai epoch pada ANN-MLR



Gambar 9. Grafik Data Prediksi dan Aktual kasus Covid-19 dengan ANN-MLR

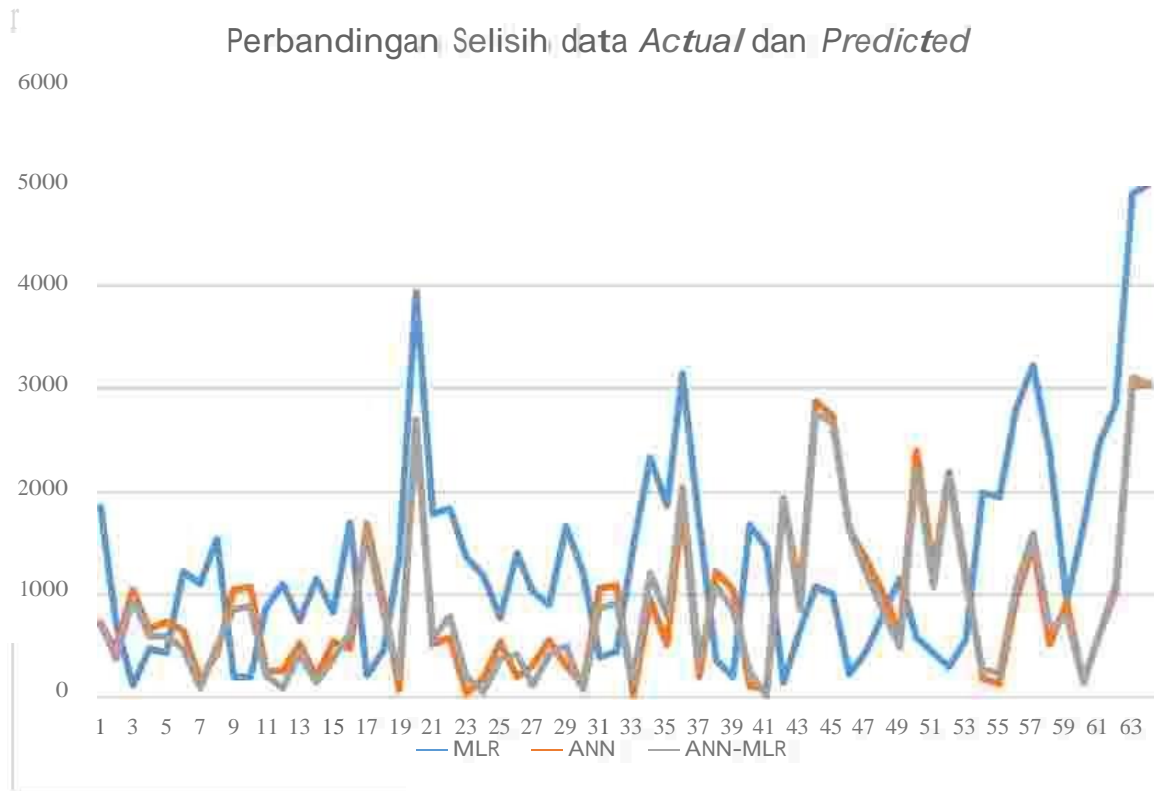
| No | Actual | Predicted | Actual-Predicted |
|----|--------|-----------|------------------|
| 1 | 5272 | 4546,16 | 725,8403 |
| 2 | 3923 | 4297,445 | 374,445 |
| 3 | 3718 | 4666,336 | 948,336 |
| 4 | 3807 | 4400,971 | 593,971 |
| 5 | 4265 | 4864,892 | 599,8916 |
| 6 | 9063 | 9534,49 | 471,49 |
| 7 | 4792 | 4882,788 | 90,7876 |
| 8 | 4998 | 4492,548 | 505,4517 |
| 9 | 4360 | 5208,079 | 848,079 |
| 10 | 4442 | 5327,738 | 885,7383 |
| 11 | 4192 | 4406,079 | 214,0786 |
| 12 | 5534 | 5629,248 | 95,248 |
| 13 | 4917 | 5332,182 | 415,182 |
| 14 | 5828 | 5965,215 | 137,2153 |
| 15 | 5418 | 5799,161 | 381,161 |
| 16 | 6267 | 5631,153 | 635,8467 |
| 17 | 4617 | 6208,313 | 1591,313 |
| 18 | 5092 | 5900,988 | 808,988 |
| 19 | 5533 | 5339,512 | 193,488 |
| 20 | 8369 | 5669 | 2700 |
| 21 | 5803 | 5235,775 | 567,2246 |
| 22 | 6027 | 5245,814 | 781,186 |
| 23 | 6089 | 5888,639 | 200,361 |
| 24 | 5754 | 5808,825 | 54,825 |
| 25 | 5292 | 5671,33 | 379,33 |
| 26 | 6058 | 5647,666 | 410,334 |
| 27 | 6033 | 6144,265 | 111,2646 |
| 28 | 6310 | 6724,873 | 414,8726 |

| | | | |
|----|-------|----------|----------|
| 29 | 6388 | 5893,61 | 494,3896 |
| 30 | 6189 | 6110,64 | 78,3604 |
| 31 | 5489 | 6352,977 | 863,9766 |
| 32 | 6120 | 7026,339 | 906,339 |
| 33 | 6725 | 6567,725 | 157,2754 |
| 34 | 7354 | 6141,088 | 1212,912 |
| 35 | 6689 | 5895,464 | 793,5356 |
| 36 | 7751 | 5705,869 | 2045,131 |
| 37 | 13830 | 14172,74 | 342,735 |
| 38 | 6347 | 7439,638 | 1092,638 |
| 39 | 4360 | 5208,079 | 848,079 |
| 40 | 7514 | 7261,758 | 252,242 |
| 41 | 7199 | 7203,494 | 4,494 |
| 42 | 7259 | 9178,265 | 1919,265 |
| 43 | 6740 | 7600,945 | 860,945 |
| 44 | 6528 | 9280,875 | 2752,875 |
| 45 | 5854 | 8524,725 | 2670,725 |
| 46 | 7903 | 9541,266 | 1638,266 |
| 47 | 8002 | 9217,656 | 1215,656 |
| 48 | 8074 | 8918,524 | 844,524 |
| 49 | 8072 | 8563,91 | 491,91 |
| 50 | 7203 | 9426,828 | 2223,828 |
| 51 | 6877 | 7950,871 | 1073,871 |
| 52 | 6753 | 8883,355 | 2130,355 |
| 53 | 7445 | 8579,23 | 1134,23 |
| 54 | 8854 | 8582,45 | 271,55 |
| 55 | 9321 | 9082,282 | 238,718 |
| 56 | 10617 | 9544,847 | 1072,153 |
| 57 | 10046 | 8460,37 | 1585,63 |
| 58 | 9640 | 8972,272 | 667,728 |
| 59 | 8692 | 9471,857 | 779,857 |
| 60 | 10047 | 10184,09 | 137,086 |
| 61 | 11278 | 10672,75 | 605,253 |
| 62 | 11557 | 10455,27 | 1101,729 |
| 63 | 12818 | 9708,666 | 3109,334 |
| 64 | 14224 | 11165,96 | 3058,041 |

Tabel 6. Nilai selisih data *Actual* dan *Predicted* pada ANN-MLR

Perbandingan Analisis Performansi

Untuk hasil performansi, hasil hybrid antara ANN dan MLR yang cukup baik, dengan nilai error dari 3 cara pengukuran yang lebih kecil dibanding dengan metode ANN dan MLR. yaitu dengan nilai RMSE 1178,68, nilai MAE 887.68 dan nilai MAPE sebesar 12.92%. Untuk perbandingan ukuran error tiap metode dapat dilihat pada Tabel 4. Perbandingan selisih antara data aktual dan prediksi pada tiap metode dapat dilihat pada Gambar 10.



Gambar 10. Grafik Perbandingan selisih Data Prediksi dan Aktual kasus Covid-19 dengan MLR, ANN, dan ANN-MLR

| UKURAN PERFORMANSI | METODE | | |
|--------------------|---------|---------|---------|
| | ANN | MLR | ANN-MLR |
| RMSE | 1202.11 | 1707.95 | 1178,68 |
| MAE | 910 | 1342.64 | 887.68 |
| MAPE | 13.51% | 18.07% | 12.92%. |

Tabel 7. Perbandingan Nilai Error Metode

5. Kesimpulan

Pada penelitian prediksi kasus Positif Covid-19 di Indonesia, metode hybrid pada ANN-MLR memberikan hasil yang lebih baik daripada ANN dan MLR yaitu dapat menurunkan nilai error pada pengukuran statistik sebesar 1,9% untuk RMSE, 2,5% untuk MAE dan 4,4% untuk MAPE. Bahkan jika dibandingkan dengan metode MLR, RMSE turun sebesar 30,1%, MAE turun sebesar 33,1% dan MAPE turun sebesar 28,6%. Dengan kesimpulan ini metode hybrid antara MLR dan ANN cukup bagus untuk mereduksi tingkat error untuk prediksi kasus Covid-19 di Indonesia. Untuk penelitian lebih lanjut dapat juga dicari atau disesuaikan jumlah parameter X atau variabel independennya untuk mendapatkan hasil yang berbeda.

Referensi

- [1] World Health Organization, "Coronavirus". 2020 [Online]. Available: <https://www.who.int/news-room/q-a-detail/q-a-coronaviruses> [diakses 21 April 2020]
- [2] Miquel, Porta. 2008 "Dictionary of Epidemiology" Oxford University Press. hlm. 179. ISBN 978-0-19-531449-6.
- [3] Badan Nasional Penanggulangan Bencana, "Laman Resmi COVID- 19" 2020. [Online]. Available: <https://bnpb-inacovid19.hub.arcgis.com/>. [Diakses 22 April 2020].
- [4] Saputra, Jimmy. 2017 "Perbandingan Model Estimasi Artificial Neural Network Optimasi Genetic Algorithm dan Regresi Linier Berganda" Statistik Indonesia
- [5] Ihwan, A. 2013. "Metode Jaringan Saraf Tiruan Propagasi Balik Untuk Estimasi Curah Hujan Bulanan di Ketapang Kalimantan Barat" Universitas Tanjungpura
- [6] Sahoo, G., Schladow, S. and Reuter, J. (2009) "Forecasting Stream Water Temperature Using Regression Analysis, Artificial Neural Network, and Chaotic Non-Linear Dynamic Models." *Journal of Hydrology*, 378, 325-342
- [7] González Vilas, L., Spyrakos, E. and Torres Palenzuela, J.M. 2011 "Neural Network Estimation of Chlorophyll a from MERIS Full Resolution Data for the Coastal Waters of Galician rias (NW Spain). *Remote Sensing of Environment*," 115, 524-535.
- [8] Dwi, Andi. 2014. "Pemanfaatan Goole Trends Dalam Penentuan Kata Kunci Sebuah Produk Untuk Meningkatkan Daya Saing Pelaku Bisnis di Dunia Internet". STIMIK AMIKOM Purwokerto
- [9] Purnaningrum, Evita. 2019. "Google Trends Analytics Dalam Bidang Pariwisata". Universitas PGRI Adi Buana Surabaya
- [10] Sidhartha S.PadhiRupesh K.Pati 2017 "Quantifying potential tourist behavior in choice of destination using Google Trends" Indian Institute of Management Kazhikode
- [11] Montano Moreno, JJ, Palmer Pol A, Muñoz Gracia P. 2011 "Artificial neural networks applied to forecasting time series"
- [12] Dar, M. H. (2017). "Penerapan Metode Backpropagation Neural Network Untuk Memprediksi Produksi Air" 12, 203– 208
- [13] Khashei, M. and Bijari, M. 2011 "A Novel Hybridization of Artificial Neural Networks and ARIMA Models for Time Series Forecasting" *Applied Soft Computing*, 11, 2664-2675.
- [14] Book 4, Hydrologic Analysis and Interpretation, Chapter A3. Statistical Methods in Water Resources. [Online] Available: <https://pubs.usgs.gov/twri/twri4a3/pdf/twri4a3-new.pdf> (Diakses 22 April 2020).
- [15] Mitra Ebrahimi a, Ali Akbar Safari Sinigani a, Mohammad Reza Sarikhani b,, dan Seyed Abolghasem Mohammadi . 2017. "Comparison Of Artificial Neural Network And Multivariate Regression Models For Prediction Of Azotobacteria Population In Soil Under Different Land Uses." Bu-Ali Sina University
- [16] Tsakiri, Katerina, Antonios M., Stelios K 2010 "Artificial Neural Network and Multiple Linear Regression for Flood Prediction in Mohawk River, New York"

- [17] Dania, Wike Agustin Prima., Shyntia Atica Putri. 2015 “Peramalan Permintaan Dengan Pendekatan Marketing Mix Pada Produk Keripik Apel Menggunakan Metode Jaringan Syaraf Tiruan (Studi Kasus Di Ukm So Kressh Malang)” Universitas Brawijaya
- [18] Iqbal, M. “An Introduction to Solar Radiation” Academic Press: Toronto, ON, Canada, 1986; pp. 215–234
- [19] Zhang, B. Eddy Patuwo, Michael Y. Hu. 1997 “Forecasting with artificial neural networks: The state of the art Guoqiang” Kent State University
- [20] Shyam J. Kurian, Atiq ur Rehman Bhatti, Mohammed Ali Alvi, Henry H. Ting, Curtis Storlie, Patrick M. Wilson, MPH; Nilay D. Shah, Hongfang Liu, Mohamad Bydon. 2020 “Correlations Between COVID-19 Cases and Google Trends Data in the United States: A State-by-State Analysis” Mayo Foundation for Medical Education and Research n Mayo Clin Proc. 2020;95(11):2370-2381
- [21] Setahun search [Online]. Available: <https://trends.google.co.id/trends/yis/2020/ID/> [diakses 10 Januari 2021]
- [22] Rawlings, J. Pantula, S.G. and Dickey, D.S. (1998) “Applied Regression Analysis. 2nd Edition” Springer-Verlag, New York.
- [23] Lola M. Safiih, Mohd Noor Afiq Ramlee, G. Sukan Gunalan, Nurul Hila Zainuddin, Razak Zakariya, MdSuffian Idris, Idham Khalil (2016) “Improved the Prediction of Multiple Linear Regression Model Performance Using the Hybrid Approach: A Case Study of Chlorophyll-a at the Offshore Kuala Terengganu, Terengganu” Open Journal of Statistics, 2016, 6, 789-804