

---

# **Analisis Dan Perbandingan Algoritma Prediksi Dalam Mengetahui Perkiraan Peningkatan Jumlah Kasus COVID-19 Di Kabupaten Boyolali Dengan Metodologi CRISP-DM**

**Erlangga Wisnu Murti**

Teknologi Informasi, Universitas Kristen Satya Wacana  
Email : [erlanggawm56@gmail.com](mailto:erlanggawm56@gmail.com)

**Abstrak :** CORONA VIRUS pertama kali dilaporkan pada tahun 2019 di Kota Wuhan China. Beberapa negara di belahan dunia yang berpotensi besar tersebarnya virus Covid-19 berusaha untuk menanggulangi penyebaran virus, dan juga termasuk Indonesia. Data dan informasi merupakan bagian penting dalam pertimbangan mengambil keputusan terkait penanganan Covid-19. Data Virus ini senantiasa dimanfaatkan untuk dilakukan analisis prediktif demi mengetahui perkiraan jumlah kasus Covid-19 di masa depan. Hasil perprediksi dapat digunakan sebagai pertimbangan bagi pemerintah dalam mengambil kebijakan dan intervensi dalam penanganan Covid-19 di Indonesia. Dalam penelitian ini akan dilakukan prediksi jumlah kasus Covid-19 dengan membandingkan beberapa model algoritma yaitu Neural Network, Random Forest, Linear Regression, Support Vector Machine, Gaussian Process, dan Polynomial Regression. Untuk mengukur tingkat akurasi dari tiap model algoritma digunakan model validasi 10 Fold Cross Validation dan evaluasi menggunakan Root Mean Square Error (RMSE) menggunakan alat bantu RapidMiner Studio v9.6.0.0. Adapun metodologi penelitian yang digunakan ialah metodologi Cross Industry Standard Process for Data Mining (CRISP-DM). Hasil penelitian ini menyimpulkan bahwa dataset kasus Covid-19 di Indonesia dapat diprediksi dengan menggunakan model algoritma Linear Regression.

**Kata Kunci:** Covid-19, CRISP-DM, Perbandingan Algoritma Prediksi, Rapid Miner

**Abstract :** CORONA VIRUS was first reported in 2019 in Wuhan City, China. Several countries in parts of the world that have a high potential for the spread of the Covid-19 virus are trying to contain the spread of the virus, and this includes Indonesia. Data and information are an important part of considering decisions regarding handling Covid-19. This virus data is always used to carry out predictive analysis to find out the estimated number of Covid-19 cases in the future. The prediction results can be used as a consideration for the government in making policies and interventions in handling Covid-19 in Indonesia. In this research, predictions will be made of the number of Covid-19 cases by comparing several algorithm models, namely Neural Network, Random Forest, Linear Regression, Support Vector Machine, Gaussian Process, and Polynomial Regression. To measure the level of accuracy of each algorithm model, the 10 Fold Cross Validation validation model was used and evaluated using the Root Mean Square Error (RMSE) using the RapidMiner Studio v9.6.0.0 tool. The research methodology used is the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology. The results of this research conclude that the dataset of Covid-19 cases in Indonesia can be predicted using the Linear Regression algorithm model

**Keywords:** Covid-19, CRISP-DM, Comparison of Prediction Algorithms, Rapid Miner

## PENDAHULUAN

Pada saat ini dunia tengah mengalami wabah virus yaitu virus “Covid-19”, penyebaran sebuah virus ini biasa dikenal dengan virus Corona. CoV (Corona Viruses) merupakan bagian dari keluarga virus yang tersebar di dunia ini saat, dan mungkin juga sudah ada pada jaman dahulu [4]. Virus tersebut dapat mengakibatkan penyakit mulai dari “Flu” hingga penyakit yang lebih berat seperti Middle East Respiratory Syndrome (MERS-CoV) dan Severe Acute Respiratory Syndrome (SARS-CoV)[2].

Organisasi Kesehatan Dunia atau (WHO) menetapkan virus corona ini sebagai pandemic 11 Maret 2020, dan Organisasi Kesehatan Dunia atau (WHO) menetapkan bahwa apabila terjangkit virus tersebut akan mengalami penyakit yang amat serius dan harus mendapatkan penanganan khusus, status pandemic atau epidemic global menandakan bahwa penyebaran COVID-19 berlangsung sangat cepat sehingga hampir belahan dunia terjangkit oleh virus tersebut, tidak ada Negara di dunia yang dapat memastikan diri dari terhindar dari virus corona, sudah terhitung 196 Negara belahan dunia yang terkonfirmasi COVID-19. Sedikit membantu menekan angka kasus di dunia maka diadakan kegiatan vaksinasi [1]. Vaksinasi sendiri termasuk penyuplaiyaan immune pada tubuh makhluk hidup, adapun 4 tujuan dan manfaat melakukan vaksinasi diantaranya : 1. Membentuk Kekebalan Kelompok (herd immunity), 2. Menurunkan kesakitan dan kematian akibat COVID-19, 3. Menekan atau memangkas kasus terjangkitnya virus corona.

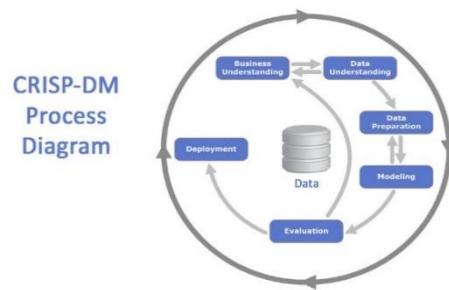
Proses Data mining dilakukan untuk keperluan memprediksi pasien yang terjangkit dan pasien yang meninggal COVID-19 di Kabupaten Boyolali. Data mining sendiri merupakan salah satu cabang Artificial Intelligence yang bekerja dengan cara mencari pola atau informasi yang menarik dengan menguraikan penemuan pengetahuan didalam database dengan menggunakan teknik-teknik tertentu [5][3].

Di Indonesia sendiri, Covid-19 merebak sejak Maret 2020 dan telah menginfeksi banyak orang di Indonesia, tercatat sampai akhir tahun 2020 mencapai 706.837 orang yang terkonfirmasi positif virus corona. Di awal tahun 2021, kasus terkonfirmasi COVID-19 terus mengalami peningkatan yaitu sebanyak 869.600 orang berdasarkan data dari Komite Penanganan COVID-19 dan Pemulihan Ekonomi Nasional [2]. Peningkatan jumlah pasien COVID-19 menjadi penting untuk dikaji lebih lanjut antara lain untuk melakukan prediksi atau perkiraan jumlah kasus COVID-19 pada waktu yang akan datang [2]

## METODE PENELITIAN

### CRISP-DM

Penjelasan alur pada penelitian ini, disesuaikan dengan metode CRISP-DM (Cross Industry Standard Process for Data Mining) menyediakan standar proses data mining sebagai strategi pemecahan masalah yang dihadapi secara umum dari Bussines atau Unit-unit peneltian.



**Gambar 1. Proses Data Mining CRISP-DM**

#### 1. Business/Research Understanding

Phase Tahap Pertama ini untuk memahami tujuan dan kebutuhan proyek secara detail dari sudut pandang bisnis ataupun unit penelitian secara keseluruhan, kemudian menerjemahkan pengetahuan ini ke dalam pendefinisian masalah pada data mining. Selanjutnya menyiapkan strategi awal untuk mencapai tujuan.

#### 2. Data Understanding

Tahap ini dimulai dengan pengumpulan data yang kemudian akan dilanjutkan dengan proses untuk menganalisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal, mengidentifikasi kualitas data untuk mendeteksi adanya bagian yang menarik dari data yang dapat digunakan untuk hipotesa untuk informasi yang tersembunyi.

#### 3. Data Preparation Phase

Tahap ini meliputi semua kegiatan untuk membangun dataset akhir (data yang akan diproses pada tahap pemodelan) dari data mentah. Tahap ini dapat diulang beberapa kali. Kemudian memilih kasus dan variabel yang ingin dianalisis dan yang sesuai analisis yang akan dilakukan, melakukan perubahan pada beberapa variable jika dibutuhkan. Siapkan data awal sehingga siap untuk perangkat permodelan

#### 4. Modeling Phase

Tahap ini akan dilakukan pemilihan dan penerapan berbagai teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang optimal. Perlu memperhatikan beberapa teknik yang mungkin digunakan untuk permasalahan data mining

yang sama. Jika diperlukan, proses dapat kembali ke fase pengolahan data kedalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik data mining tertentu.

### 5. Evaluation Phase

Tahap ini akan dilakukan pemilihan dan penerapan berbagai teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang optimal. Perlu memperhatikan beberapa teknik yang mungkin digunakan untuk permasalahan data mining yang sama. Jika diperlukan, proses dapat kembali ke fase pengolahan data kedalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik data mining tertentu.

### 6. Deployment Phase

Pada tahap ini, pengetahuan atau informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh pengguna

## HASIL PEMBAHASAN

Implementasi data mining dan analisa data pada penelitian ini terdiri dari beberapa tahapan berdasarkan konsep metodologi CRISP-DM.

### A. Pemahaman Bisnis (Business Understanding)

Pada tahap ini diperlukan pemahaman tentang latar belakang dan tujuan pada proses bisnis yang berhubungan dengan perkembangan covid19 di Kab.Boyolali, diantaranya :

- Menentukan Tujuan Bisnis (Determine Business Objectives)
- Menilai Situasi (Assess the Situation)
- Menentukan Tujuan Data Mining (Determine the Data Mining Goals)

### B. Pemahaman Data (Data Understanding)

Tahapan pemahaman data diawali dengan pengumpulan data awal atau data mentah dan akan melakukan pengolahan data, hasil kegiatan ini dalam rangka untuk membiasakan diri dengan data guna untuk membantu mengidentifikasi masalah data.

- Mengumpulkan Data Awal (Collect the Initial Data)

Pengumpulan data dilakukan dengan mengajukan permintaan dataset covid-19 yang bersifat private yang ada di Dinas Kesehatan (DINKES) Kab.Boyolali. Dataset covid-19 tersebut yang digunakan bersumber langsung dari Dinas Kesehatan (DINKES) Kab.Boyolali

The table consists of two side-by-side Microsoft Excel spreadsheets. The left spreadsheet has columns A through I and rows 1 through 45. The right spreadsheet has columns A through I and rows 09/01/2021 through 04/02/2021. Both spreadsheets contain data in a grid format with various formulas and functions applied to the first few rows.

**Tabel 1. Dataset Mentah Kasus Covid-19 di Kabupaten Boyolali Periode Januari 2021 – Desember 2021**

- Normalisasi Data & Mendeskripsikan Data

Normalisasi Data dilakukan bertujuan untuk mengurangi dan mencari dataset yang tidak valid, menghilangkan data-data kosong, dan mengetahui missing value.

A	B	C	D
Date	Tolal Kasus	Kasus Meninggal	Kasus Sembuh
January 1, 2021	1	0	1
January 2, 2021	2	0	2
January 3, 2021	9	0	7
January 4, 2021	13	1	7
January 5, 2021	40	1	39
January 6, 2021	53	2	52
January 7, 2021	61	3	61
January 8, 2021	66	4	63
January 9, 2021	74	6	70
January 10, 2021	186	8	178
12			
A	B	C	D
Date	Tolal Kasus	Kasus Meninggal	Kasus Sembuh
March 1, 2021	32	2	30
March 2, 2021	32	0	32
March 3, 2021	13	0	13
March 4, 2021	25	2	23
March 5, 2021	9	1	8
March 6, 2021	8	0	8
March 7, 2021	3	0	3
March 8, 2021	32	2	30
March 9, 2021	0	0	0
March 10, 2021	27	1	26
24			
F	G	H	I
Date	Tolal Kasus	Kasus Meninggal	Kasus Sembuh
February 1, 2021	93	7	86
February 2, 2021	48	5	43
February 3, 2021	50	1	49
February 4, 2021	35	1	34
February 5, 2021	27	1	26
February 6, 2021	16	2	14
February 7, 2021	4	0	4
February 8, 2021	37	0	37
February 9, 2021	0	0	0
February 10, 2021	63	0	63
24			
F	G	H	I
Date	Tolal Kasus	Kasus Meninggal	Kasus Sembuh
April 1, 2021	20	1	19
April 2, 2021	0	0	0
April 3, 2021	12	1	11
April 4, 2021	1	1	0
April 5, 2021	37	1	36
April 6, 2021	36	2	34
April 7, 2021	27	1	26
April 8, 2021	20	1	19
April 9, 2021	0	0	0
April 10, 2021	4	2	2

**Tabel 2. Dataset Hasil Normalisasi dari data mentah awal**

- Mendeskripsikan Data (Describe the Data) Dataset yang didapatkan menggunakan format excel dari tanggal 1 Januari 2021 sampai dengan 10 Desember 2021 yang berjumlah 7497 ribu data serta memiliki atau diambil 4 Atribut yang akan diambil.
- Mengeksplorasi Data (Explore Data)

Atribut	Keterangan
Tanggal	Menjelaskan tentang data hasil observasi per hari oleh DINKES

Total Kasus	Jumlah total kasus terkonfirmasi Covid-19 di Kabupaten Boyolali Jawa Tengah
Kasus Meninggal	Jumlah total kasus meninggal di Kabupaten Boyolali per hari
Kasus Sembuh	Jumlah total kasus sembuh di Kabupaten Boyolali per hari

**Tabel 3. Deskripsi Atribut**

Atribut	Keterangan
Tanggal	Date
Total Kasus	Numeric
Kasus Meninggal	Numeric
Kasus Sembuh	Numeric

**Tabel 4. Explorasi Data**

### C. Persiapan Data (Data Preparation)

Data yang digunakan akan diolah dengan menggunakan tools RapidMiner Studio x64 9.10.8. Tahapan persiapan data diambil dari kegiatan hasil normalisasi terakhir dari dataset awal atau mentah. Data hasil normalisasi tersebut yang akan dimasukkan ke dalam pemodelan (RapidMinerStudio). Ada beberapa tahap yang dilakukan pada persiapan data yaitu menentukan data yang dibutuhkan, pengecekan missing value, serta mengambil atribut data yang akan dimasukkan ke dalam pemodelan. Dataset yang akan digunakan yaitu dataset hasil normalisasi atau dataset pada Tabel 2. Atribut data yang dipilih guna untuk diolah terdiri 4 atribut Date, Total Kasus, Kasus Meninggal, Kasus Sembuh. Atribut Date yang akan dijadikan sebuah “Label/Id”. Sedangkan atribut yang lain ialah numeric.

The image shows two side-by-side screenshots of the RapidMiner Studio 'Import Data' interface. The left screenshot is titled 'Select the cells to import' and shows a table with four columns: A, B, C, and D. The data rows include columns for Date, Total Kasus, Kasus Meninggal, and Kasus Sembuh. The right screenshot is titled 'Format your columns' and shows the same data with the columns labeled Date, Total Kasus, Kasus Meninggal, and Kasus Sembuh. The 'Date' column is highlighted as the label. Both screenshots show a list of data rows from Jan 1, 2021, to Feb 2, 2021.

**Gambar 2. Persiapan Data & Menentukan Label**

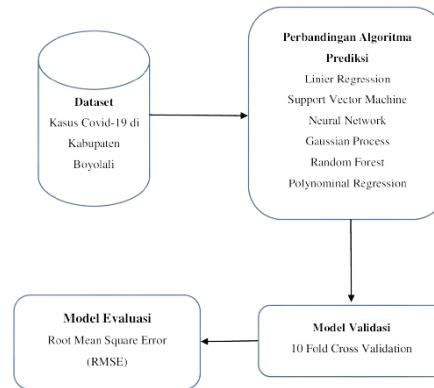
## D. Pemodelan (Modeling)

Masuk pada tahapan pemodelan, ada beberapa hal yang akan dilakukan antara lain menentukan teknik pemodelan, membangun model, dan menilai model.

- Menentukan Teknik Pemodelan

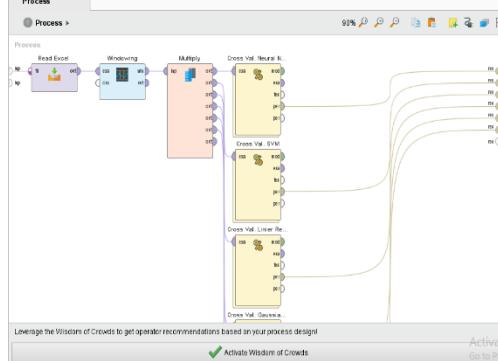
Untuk menentukan pemodelan dilakukan dengan melakukan perbandingan dari beberapa algoritma-algoritma prediksi. Langkah-langkah untuk melakukan perbandingan ada pada gambar 3 yang terdiri dari :

- Dataset
- Algoritma-algoritma Prediksi
- Model Validasi
- Model Evaluasi



**Gambar 3. Langkah-langkah perbandingan algoritma prediksi**

Langkah selanjutnya mengimport dataset kasus covid-19 ke RapidMiner Studio. Setelah import dataset, beserta tahapan persiapan data, dilakukan pengukuran dan perbandingan algoritma prediksi dengan pembentukan proses learning dan juga testing pada setiap model beberapa algoritma



**Gambar 4. Proses Learning dan Testing pada setiap model algoritma**

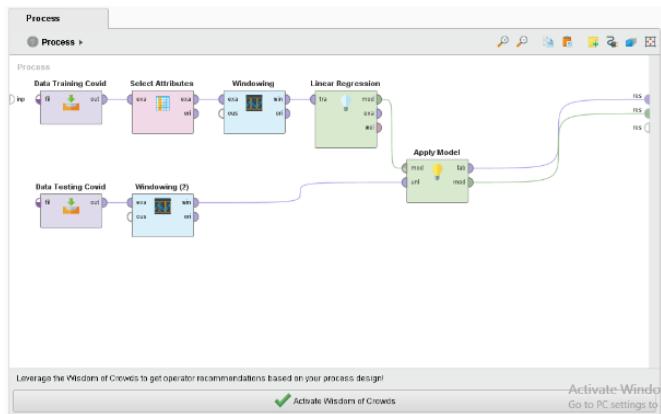
Algoritma	RMSE
<i>Linear Regression</i>	72.786 +/- 34.972 (micro average: 79.126 +/- 0.000)
<i>Random Fore</i>	84.613 +/- 49.805 (micro average: 96.682 +/- 0.000)
<i>Neural Network</i>	97.557 +/- 40.239 (micro average: 104.723 +/- 0.000)
<i>SVM</i>	109.674 +/- 46.323 (micro average: 118.459 +/- 0.000)
<i>Gaussian Process</i>	119.575 +/- 85.939 (micro average: 143.155 +/- 0.000)
<i>Polynomial Regression</i>	135.759.551 +/- 419.199.674 (micro average: 410.984824 +/- 0.000)

**Tabel 5. Hasil RMSE Setiap Algoritma**

Dilihat dari hasil pengujian pada setiap algoritma prediksi menunjukkan bahwa algoritma yang memiliki RMSE terkecil ialah Linier Regression dengan hasil akurasi prediksi RMSE 72.786 +/- 34.972 (micro average) 79.126 +/- 0.000), jadi dapat disimpulkan bahwa algoritma yang akan digunakan untuk memprediksi jumlah kasus covid-19 di Kab.Boyolali adalah Linier Regression

- Build Model (Membangun Model)

Menampilkan skema model yang dibentuk untuk memprediksi kasus covid-19 di Indonesia menggunakan algoritma Linier Regression



**Gambar 5. Model Prediksi Kasus Covid-19 dengan Linier Regression**

Hasil dari penelitian menyimpulkan bahwa data kasus covid-19 di Kab. Boyolali dapat diprediksi dengan menggunakan model algoritma Linier Regression dengan hasil pengujian tabel hasil prediksi total kasus Covid-19 perhari periode 1 Januari 2021 sampai dengan 10 Desember 2021

Result History | LinearRegression (Linear Regression) | ExampleSet (Apply Model)

Open in: Turbo Prep | Auto Model | Filter (115 / 115 examples): all

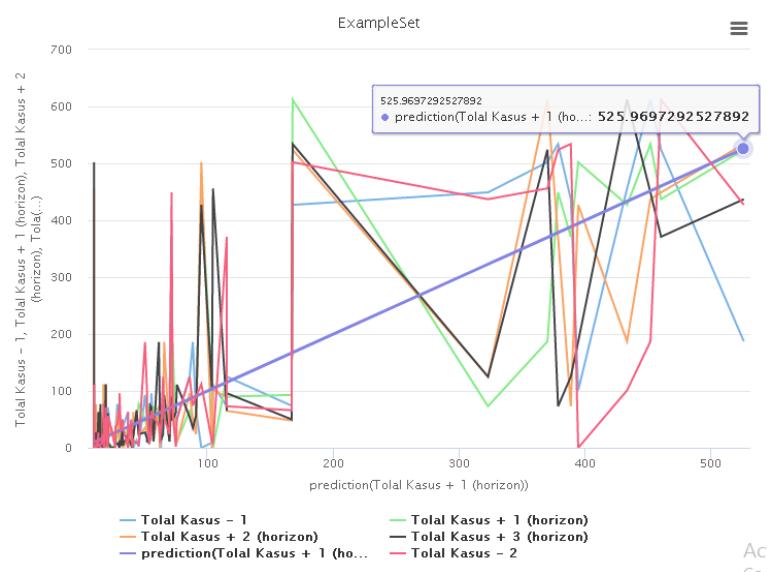
Data | Statistics | Visualizations | Annotations

Row No.	Last Date in ...	prediction(T...)	Total Kasus ...					
1	Jan 3, 2021 1...	17.415	13	40	53	1	2	9
2	Jan 4, 2021 1...	20.788	40	63	81	2	8	13
3	Jan 5, 2021 1...	43.598	53	61	66	9	13	40
4	Jan 6, 2021 1...	54.523	61	66	74	13	40	53
5	Jan 7, 2021 1...	81.270	66	74	188	40	53	81
6	Jan 8, 2021 1...	65.487	74	186	93	53	61	56
7	Jan 9, 2021 1...	72.234	186	63	48	61	66	74
8	Jan 10, 2021 ...	168.882	93	48	50	66	74	186
9	Feb 1, 2021 1...	88.258	48	50	35	74	186	93
10	Feb 2, 2021 1...	50.306	50	35	27	186	93	48
11	Feb 3, 2021 1...	51.993	35	27	16	93	48	50
12	Feb 4, 2021 1...	38.343	27	16	4	48	50	35
13	Feb 5, 2021 1...	32.596	18	4	37	50	35	27
14	Feb 6, 2021 1...	23.318	4	37	0	35	27	16
15	Feb 7, 2021 1...	13.198	0	0	63	27	16	4

ExampleSet (115 examples, 5 special attributes, 3 regular attributes)

**Gambar 6. Hasil Prediksi Total Kasus Covid-19**

Selanjutnya, berdasarkan grafik yang ditunjukkan oleh gambar 8 di bawah ini, hasil prediksi jumlah kasus Covid-19 Kab.Boyolali pada saat itu memiliki pola yang cenderung menaik. Dan pola pergerakan data prediksi dengan data aktualnya sama, serta gap antara data aktual dan data prediksi total kasus Covid-19 tidak terlalu jauh berbeda, sehingga hasil prediksi bisa dikatakan cukup baik



**Gambar 7. Grafik Prediksi Laju Perkembangan Kasus Covid-19 di Kabupaten Boyolali Periode 1 Januari 2021 – 1 Desember 2021**

## **KESIMPULAN**

Berdasarkan penelitian yang dilakukan oleh penulis yang bertujuan untuk melakukan analisis data Kasus Covid-19 dan melakukan usaha prediksi mendasar dengan melakukan perbandingan dari berbagai Algoritma Prediksi dari data mining dengan Algoritma yang digunakan ialah “Linier Regression” pada dataset kasus Covid-19 diperoleh nilai Root Mean Square Error (RMSE) sebesar : 72.786 +/- 34.972 (micro average: 79.126 +/- 0.000), merupakan RMSE yang memiliki nilai paling kecil diantara hasil Algoritma lainnya. Penerapan metodologi CRISP-DM dapat membantu melakukan Prediksi Kasus Covid-19 di Kabupaten Boyolali Jawa Tengah

## **SARAN**

Saran yang dapat diberikan untuk pengembangan lebih lanjut dalam mencakup kemungkinan adanya gelombang kedua dari COVID-19, atau tren kasus positif yang terus memuncak di waktu dekat ini, maka dilakukan pengembangan dan memprediksi yang berkelanjutan pada waktu yang akan datang. Selain mempelajari dan mempertimbangkan berbagai metode riset terkini terkait COVID-19 ialah penting untuk selalu membandingkannya dengan situasi yang terjadi di lapangan. Dengan demikian, berbagai asumsi yang ada dalam tatanan riset/kajian dapat selalu dibandingkan dengan kondisi nyata yang terjadi, dan juga dapat mengeksplor penelitian lebih luas lagi menggunakan Algoritma-algoritma yang lain

## **DAFTAR PUSTAKA**

- A. N. A. L. P. Naomi Yulia Indah Ginting, "Estimasi Magnitudo Gempa Bumi Dari Sinyal Seismik Gelombang P Menggunakan Metode Regresi Polinominal," e- Proceeding of Engineering, vol. 7, p. 4635, 2020
- A. Karim, "Perbandingan Prediksi Kemiskinan di Indonesia Menggunakan Support Vector Machine (SVM) dengan Regresi Linear," Jurnal Sains Matematika dan Statistika, vol. 6, p. 108=109, 2020
- A. P. B. D. M. Alvina Felicia Watratan, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid 19 Di Indonesia," Journal Of Applied Computer SCIENCE AND TECHNOLOGY (JACOST), vol. 1, p. 7, 2020
- A. P. Fadillah, "Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ)," Jurnal Teknik Informatika dan Sistem Informasi, vol. 1, 2015
- D. T. W. P. D. K. Ardhy Chaeruna Salim, "Prediksi Jumlah Penderita Demam Berdarah Dengue Di Kota Bandung Menggunakan Support Vector Regression Dan Gaussian Process Regression," vol. 7, p. 4915, 2020

- N. Yanti, "Penerapan Metode Neural Network Dengan Struktur Backpropagation Untuk Prediksi Stok Obat Di Apotek (Studi kasus: Apotek ABC)," Seminar Nasional Aplikasi Teknologi Informasi
- P. K. Kesehatan, "Peta Sebaran Covid-19, " Satgas Tugas Penanggana Covid-19, 01 2020. [Online]. Available: <https://covid19.go.id/peta-sebaran-covid19>. [Accessed 15 01 2020]
- S. E. Damayanti, "SWASTA, Analisis dan Implementasi Framework CRISP-DM (CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING) UNTUK CLUSTERING PERGURUAN TINGGI"
- S. W. W. R. Fahmi Ardiansyah "Data Mining Berdasarkan Analisis Runtun Waktu Untuk Pembuatan Model Prediksi Pasien Terjangkit COVID-19 dan Pasien Meninggal karena COVID-19 di Indonesia," 2020
- T.I.E.R. Tri Indarwati, "Penggunaan Metode Linier Regression Untuk Prediksi Penjualan Smartphone," Jurnal TIKomSIN, vol. 6, p. 108=109, 2020
- U. I. M. R. R. L. S. D. Agus Perdana Windarto, "Bagian 1: Kombinasi Metode Klastering dan Klasifikasi (Kasus Pandemi Covid-19 di Indonesia)," Jurnal Media Informatika Budidarma, vol. 4, 2020