

Prediksi Penduduk Miskin di Daerah Tertinggal Indonesia dengan Algoritma Prophet

Budi Wijaya Rauf

Sistem dan Teknologi Informasi, Universitas Muhammadiyah
Kendari

Email: budi.wijaya@umkendari.ac.id

Naskah diserahkan: 31-07-2023;
Direvisi: 07-08-2023;
Diterima: 10-08-2023.

ABSTRAK: Kemiskinan menjadi salah satu masalah umum yang dihadapi oleh setiap negara di dunia, termasuk Indonesia. Pemerintah dari berbagai daerah yang ada di Indonesia sedang berjuang untuk memberantas kemiskinan, begitu pun dengan daerah-daerah tertinggal. Tercatat di Badan Pusat Statistik (BPS) bahwa terjadi kenaikan angka kemiskinan dari 18,87% (2015) kini menjadi 24,56% (2022) di daerah-daerah tertinggal yang terdiri dari 62 daerah. Dengan memprediksi jumlah penduduk miskin di daerah tertinggal, diharapkan pemerintah bisa mengambil langkah untuk memberantas kemiskinan. Penelitian ini menggunakan algoritma *Prophet* yang selanjutnya dibangun ke dalam pemodelan *machine learning* dalam memprediksikan penduduk miskin di daerah tertinggal. Data yang digunakan adalah data dari tahun 2015-2022, tujuan dari penelitian ini adalah untuk melihat tingkat akurasi dan *error* dengan menggunakan parameter *Mean Squared Error (MSE)* dan *Mean Absolute Percentage Error (MAPE)*. Hasil dari prediksi menunjukkan kenaikan masyarakat miskin menjadi 35% pada tahun 2024 dan diprediksikan akan terus naik hingga 35% pada tahun 2027 dengan parameter MSE sebesar 2% dan MAPE sebesar 1%.

Katakunci: Daerah Tertinggal, *Forecasting*, Kemiskinan, *Machine Learning*, *Prophet*

ABSTRACT: Poverty is one of the common problems faced by every country in the world, including Indonesia. Governments from various regions in Indonesia are striving to eradicate poverty, particularly in underdeveloped areas. According to the Central Bureau of Statistics (BPS), the poverty rate has increased from 18.87% in 2015 to 24.56% in 2022 in these underdeveloped areas, which consist of 62 regions. By predicting the number of poor people in these regions, the government hopes to take steps to combat poverty. This study employs the Prophet algorithm, which is then integrated into a machine learning model to predict the number of poor people in underdeveloped areas. The data used covers the years 2015 to 2022, and the aim of this research is to assess the accuracy and error rate using the Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) parameters. The prediction results indicate an increase in the number of impoverished individuals to 35% in the year 2024, and it is projected to continue rising to 35% in 2027, with an MSE parameter of 2% and an MAPE parameter of 1%.

Keywords: Underdeveloped regions; *Forecasting*; Poverty; *Machine Learning*; prophet

PENDAHULUAN

Kemiskinan menjadi salah satu permasalahan utama yang ada di berbagai belahan dunia. Setiap negara berbondong-bondong untuk memberantas kemiskinan agar mencapai cita-cita setiap bangsa yaitu menjadi bangsa yang makmur dan sejahtera, Indonesia pun tidak luput akan hal ini (Pramunendar, Dewi, & Asari, 2013). Kemiskinan merupakan individu atau kelompok masyarakat yang tidak memiliki kemampuan untuk memenuhi kebutuhan dasar. Konsep kemiskinan ini direferensikan dari buku yang dikeluarkan oleh Bank Dunia "*Handbook on Poverty and Inequality*" yang menjelaskan bahwa ketidakmampuan seorang individu secara ekonomi untuk memenuhi kebutuhan dasar makanan. Selanjutnya, penduduk akan dikategorikan sebagai penduduk miskin jika memiliki pengeluaran rata-rata per kapita per bulan dibawah garis kemiskinan.

Garis kemiskinan (GK) merupakan nilai terendah yang diperlukan seorang individu untuk memenuhi kebutuhan pokok hidupnya selama satu bulan, baik untuk kebutuhan makanan maupun non-makanan. Oleh karena itu, GK terdiri menjadi dua bagian yaitu Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non-Makanan (GKNM). Upaya untuk memberantas kemiskinan bukan hanya dilakukan oleh pihak pemerintahan, tetapi juga dari berbagai elemen masyarakat. Penelitian yang dilakukan oleh Syaharuddin, dkk (2020) menggunakan algoritma *ANN Back Propagation* untuk memprediksikan angka kemiskinan yang ada di Indonesia yang menunjukkan hasil bahwa sekitar 332.005 jiwa masih berada dibawah garis kemiskinan.

Menurut Badan Pusat Statistik (BPS) terdapat 25 juta masyarakat miskin di Indonesia dan sekitar 24,56% (2022) diantaranya berasal dari daerah tertinggal. Daerah tertinggal termuat dalam Peraturan Presiden (PP) Nomor 131 Tahun 2015 tentang Penerapan Daerah Tertinggal Tahun 2015-2019 (122 daerah tertinggal). Pada tahun 2020, berdasarkan Peraturan Presiden Nomor 63 Tahun 2020 tentang Penetapan Daerah Tertinggal Tahun 2020-2024 (62 Daerah Tertinggal).

Metode Penelitian

Pendekatan penelitian ini adalah kuantitatif deskriptif. Data yang digunakan pada penelitian ini merupakan data *time series* dalam kurun waktu 8 tahun, mulai dari tahun 2015 hingga tahun 2022. Data tersebut diambil dari website BPS, data yang menjadi acuan dalam penggunaan algoritma *prophet* adalah persentase masyarakat miskin di daerah tertinggal dengan variable yang bersifat dependen.

Tabel 1 Data Persentase Kemiskinan Di Daerah Tertinggal Tahun 2015-2022 (BPS, 2023)

No	Tahun	Persentase Kemiskinan (%)
1	2022	24.56
2	2021	25.5
3	2020	25.32

4	2019	17.01
5	2018	17.39
6	2017	18.07
7	2016	18.43
8	2015	18.87

Prophet merupakan sebuah algoritma dengan basis *additive model* yang berarti data yang dibutuhkan memiliki variable yang dependen, algoritma *prophet* juga memiliki kemampuan permodelan yang baik dalam mengenali pola musiman. *Library* yang digunakan untuk perancangan model algoritma *prophet* diantara lain *pandas*, *numpy*, *matplotlib*, *prophet*, *prophet diagnostic*, dan *prophet plot*. Dalam membangun model *machine learning* dengan algoritma *prophet*, langkah-langkah yang harus dilakukan adalah pengambilan data, pra proses data, prediksi dan validasi.

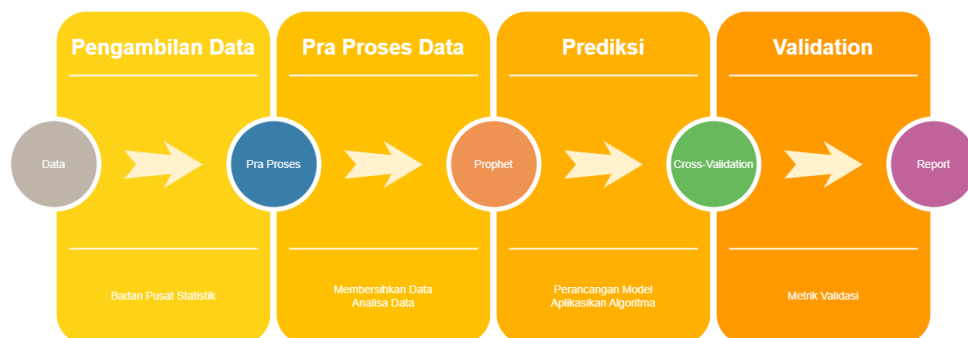
Langkah pengambilan data dilakukan dengan mengakses *website* BPS dan mengunduh data presentase kemiskinan di daerah tertinggal tahun 2015-2022 dalam bentuk *csv*. Data yang telah diambil, akan melewati pra proses data, tahapan pra proses data ini memiliki beberapa langkah di dalamnya, seperti langkah pembersihan data, integrasi data, dan transformasi data.

Langkah berikutnya adalah tahap prediksi. Pada tahapan ini, model *machine learning* dengan menggunakan *library python* algoritma *prophet* akan dibangun. Tahapan prediksi mencakup membangun model, memasukan data yang telah melewati tahapan pra proses ke dalam model yang telah dibangun, lalu memprediksikan data berdasarkan algoritma *prophet*. Terakhir, data yang telah diprediksikan akan melewati uji validasi dengan menggunakan parameter yang dapat melihat tingkat akurasi dan *error* dari algoritma *prophet*. Parameter yang digunakan dalam mengukur keakuratan algoritma ini adalah *Mean Squared Error (MSE)* dan *Mean Absolute Percentage Error (MAPE)*, untuk kedua parameter tersebut, semakin kecil nilainya maka semakin baik pula model prediksi yang dibangun.

HASIL PENELITIAN DAN PEMBAHASAN

Hasil Penelitian

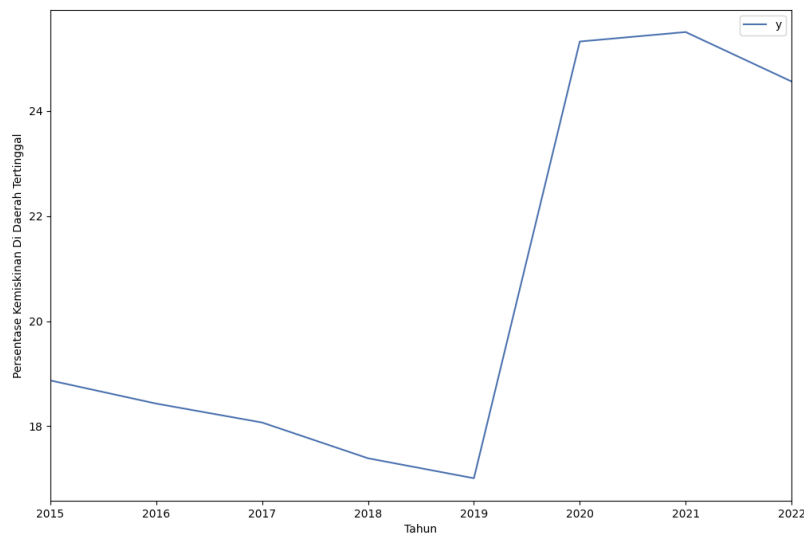
Proses penelitian ini melewati beberapa tahapan yaitu, Pengambilan Data, Pra Proses Data, Prediksi, dan Validasi.



Gambar 1. Flowchart Penelitian

Pengambilan Data

Data yang digunakan diambil dari *website* Badan Pusat Statistik <https://www.bps.go.id/> dimana data yang didapatkan berisi 2 kolom dan 8 baris data dengan rentang tahun 2015 – 2022 seperti pada tabel 1. Dalam data tersebut terdapat persentase total dari seluruh total penduduk warga miskin dari daerah tertinggal. Dari data tersebut, didapatkan fakta bahwa terjadi kenaikan penduduk miskin yang awalnya berada pada 18,87% (2015) kini telah menjadi 24,56 % (2022). *File* yang diunduh dalam bentuk *csv* yang memungkinkan pembacaan data bisa lebih cepat ketika diproses memakai bahasa pemrograman *Python*. Data dari BPS bisa dilihat pada Gambar 2 berikut ini.



Gambar 2. Grafik Kemiskinan Dalam Kurun Waktu 2015 – 2022

Pra Proses Data

Setelah melewati proses pengambilan data, tahap selanjutnya adalah *pre-processing data* (pra proses data). Tahap ini mencakup membersihkan data, integrasi data dan transformasi data (Aurelien, 2017). Selain itu, data akan dianalisis untuk digunakan dalam permodelan algoritma *prophet*.

Pembersihan data memiliki beberapa proses tersendiri. Salah satu proses yang dilakukan pada penelitian ini adalah untuk melihat jumlah data kosong pada dataset yang digunakan.

```
[7] data.isnull().sum()

tahun;kemiskinan    0
dtype: int64
```

Gambar 3. Mengecek Jumlah Data Yang Kosong

Dari gambar 3, dapat dilihat bahwa data yang telah didapatkan tidak memiliki data yang kosong. Langkah selanjutnya adalah integrasi data. Tahap integrasi data merupakan proses menjaga dataset yang berasal dari berbagai sumber agar dapat memiliki format yang sama.

Setelah memastikan data dalam bentuk format yang sama, tahap berikutnya adalah transformasi data. Data yang berasal dari Badan Pusat Statistik tidak memiliki format tipe data yang sesuai.

```
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8 entries, 0 to 7
Data columns (total 1 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   tahun;kemiskinan      8 non-null      object
dtypes: object(1)
memory usage: 192.0+ bytes
```

Gambar 4. Jumlah Kolom dan Tipe Data

Pada Gambar 4 diatas terlihat jelas bahwa jumlah kolom hanya ada satu saja, sedangkan kebutuhan analisis mengharuskan kolom berjumlah lebih dari 1. Maka dari itu, langkah selanjutnya adalah memecah kolom “tahun;kemiskinan” menjadi dua kolom. Fungsi *split()* dengan *delimiter “;”* digunakan untuk memecah kolom tersebut menjadi dua. Lalu, agar data yang digunakan bisa dianalisis oleh model algoritma *prophet*, kolom tahun diubah menjadi kolom *ds* dan kolom kemiskinan diubah menjadi kolom *y* seperti pada gambar 5 berikut ini.

```
datad.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8 entries, 0 to 7
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype
---  ---
0   ds      8 non-null      object
1   y       8 non-null      object
dtypes: object(2)
memory usage: 256.0+ bytes
```

Gambar 5. Dua Kolom Baru Yaitu Kolom *ds* dan Kolom *y*

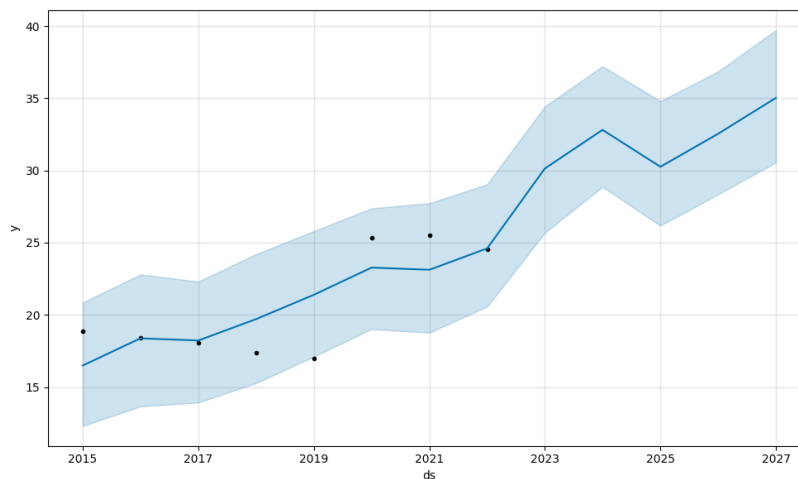
Pada Gambar 5 terlihat bahwa tipe data untuk kolom *ds* masih dalam bentuk *object*, sedangkan untuk kebutuhan analisis permodelan mengharuskan kolom tersebut dalam bentuk *datetime64*. Begitu pun juga dengan tipe data kolom *y* akan diubah menjadi tipe data *float64*.

Prediksi

Setelah melewati tahap pra proses data, tahap selanjutnya adalah merancang model prediksi menggunakan algoritma *prophet*. Perancangan model algoritma *prophet* menggunakan *library prophet* yang bisa digunakan dalam bahasa pemrograman *python*.

Langkah pertama dalam merancang model prediksi menggunakan algoritma *prophet* adalah memanggil fungsi *Prophet()* yang sebelumnya telah di-*import* ke dalam *Integrated Development Environment* (IDE). Setelah itu, data yang telah melewati pra proses dimasukkan ke dalam fungsi yang telah dipanggil dengan menggunakan *syntax*: *model.fit(data)*.

Langkah selanjutnya, membuat variabel baru yang bernama *future* yang memuat frekuensi waktu prediksi, dalam penelitian ini akan dilakukan prediksi selama 5 tahun ke depan. Lalu, variabel *future* digunakan untuk fungsi *predict()* yang berfungsi untuk memprediksikan sesuai dengan frekuensi waktu yang telah termuat dalam variabel *future*. Setelah itu, fungsi *predict()* akan dimasukkan ke dalam variabel *forecast* yang mana akan memuat data hasil prediksi dari algoritma *prophet*. Setelah langkah-langkah diatas dilakukan, maka hasilnya pun dapat terlihat. Hasil prediksi menggunakan algoritma *prophet* menunjukkan bahwa pada tahun 2027, angka kemiskinan naik menjadi 35%. Hasil prediksi dapat dilihat pada Gambar 6.



Gambar 6. Grafik Hasil Prediksi

Kolom hasil data prediksi dari variabel *forecast* terdiri dari 4 kolom yaitu *ds* yang merupakan kolom tanggal, *yhat* merupakan kolom nilai rata-rata prediksi, *yhat_lower* merupakan nilai terendah hasil prediksi, dan *yhat_upper* merupakan nilai tertinggi dari hasil prediksi. Untuk nilai dari hasil prediksi dapat dilihat pada Tabel 2 berikut ini.

Tabel 2. Hasil Prediksi

No	ds	yhat	Yhat lower	Yhat upper
1	2022-12-31	30.13	25.56	34.44

2	2023-12-31	31.81	28.86	37.22
3	2024-12-31	30.26	26.17	34.81
4	2025-12-31	32.54	28.34	36.86
5	2026-12-31	35.03	30.55	39.72

Validasi

Validasi merupakan cara untuk mengetahui apakah data hasil prediksi dari sebuah permodelan *machine learning* dapat dipercaya atau tidak. Parameter kepercayaan dapat diukur dari tingkat akurasi yang dimiliki oleh sebuah model prediksi, semakin tinggi akurasinya maka akan semakin baik. Selain itu, untuk melihat sebuah model prediksi *machine learning* yang baik dapat dilihat dari nilai *MAE* dan *MAPE*, semakin rendah nilai dari kedua parameter tersebut maka semakin baik model prediksi yang telah dibentuk.

Dalam melakukan uji validasi, penelitian ini menggunakan metode *Cross-Validation* (CV). Dengan metode tersebut, setiap data memiliki kesempatan yang sama untuk dijadikan sebagai data uji dan data latih. Untuk melakukan uji validasi, dapat menggunakan *library prophet diagnostic* yang mana telah tersedia di dalam *package library* dari *prophet*. Pengujian validasi prediksi dilakukan sebanyak 9 kali dengan data inisial yang digunakan untuk melatih *machine learning* sebanyak 730 hari, periode prediksi sebanyak 180 hari, dan horizon sebanyak 365 hari. Hasil uji validasi menggunakan metode CV menghasilkan nilai *MSE* sebesar 2% dan nilai *MAPE* sebesar 1%.

Pembahasan

Hasil dari prediksi menggunakan model *machine learning* dengan algoritma *prophet* menunjukkan bahwa presentase kemiskinan pada daerah-daerah tertinggal di Indonesia akan mengalami peningkatan dengan presentase pada tahun 2027 berada pada angka 35.03%. Nilai dari prediksi tersebut telah melewati uji validasi dan menghasilkan nilai *MSE* sebesar 2% dan *MAPE* sebesar 1%, artinya hasil prediksi dengan algoritma *prophet* ini dapat dijadikan acuan oleh pemerintah, baik pemerintah pusat maupun daerah tertinggal, dalam perencanaan pengentasan kemiskinan.

Meninjau perbandingan dari algoritma dalam memprediksikan kemiskinan di Indonesia, Abdul Karim (2020) melakukan penelitian untuk membandingkan algoritma *support vector machine* dan regresi linear. Penelitian tersebut menunjukkan bahwa nilai *MSE* dari SVM jauh lebih kecil daripada regresi linear. Selain itu, Bakti (2012) melakukan prediksi dan interpolasi di Provinsi Jawa Timur dengan metode gaussian, exponential dan Spherical dengan hasil analisis uji moran I terdapat auto korelasi spasial pada presentase jumlah penduduk miskin.

Memprediksikan jumlah kemiskinan pada kabupaten/kota di Provinsi Riau juga pernah dilakukan oleh Wanto, A. (2018) yang menerapkan jaringan saraf tiruan. Data yang digunakan merupakan data dari BPS dari tahun 2010 hingga 2015. Penelitian ini menggunakan algoritma *backpropagation* dan menghasilkan tingkat akurasi hingga 92% dengan menerapkan 5 model arsitektur. Lalu, dalam

penelitian yang dilakukan oleh Alhkarif, S., et al (2018) menganalisa *time series* dan memprediksikan harga dari *cloud computing* menggunakan algoritma *prophet* dan menghasilkan bahwa model algoritma tersebut lebih akurat dalam memprediksikan harga dari *cloud computing* untuk beberapa vendor seperti Amazon.

Pengambilan keputusan sangat bergantung dengan data yang digunakan. Semakin baik data yang digunakan, maka semakin baik pula keputusan yang dibuat. Tingkat akurasi dari model prediksi pun akan sangat mempengaruhi dalam pengambilan keputusan, dengan tingkat akurasi yang tinggi maka data yang dihasilkan pun akan semakin baik. Penelitian yang dilakukan oleh Ye Ziyuan (2019) menggunakan algoritma *ARIMA* dan *prophet* dalam memprediksikan polusi udara di Shenzhen, hasilnya menunjukkan bahwa metode hybrid tersebut dapat meningkatkan tingkat akurasi dari prediksi polusi udara. Selanjutnya, penelitian yang dilakukan oleh Rahimi, et al (2021) yaitu menganalisis dan memprediksikan kasus COVID-19 di Australia, Italia dan Inggris. Hasilnya adalah algoritma *prophet* memiliki performa prediksi yang lebih baik dibandingkan dengan model SIR dan SEIQR. Hasil penelitian tersebut lalu digunakan dalam perencanaan menanggulangi COVID-19 yang ada di Australia, Italia dan Inggris.

SIMPULAN

Berdasarkan hasil penelitian, maka diperoleh kesimpulan bahwa algoritma *prophet* yang digunakan untuk memprediksikan persentase penduduk miskin yang ada di daerah tertinggal di Indonesia berdasarkan Peraturan Presiden Nomor 63 Tahun 2020 tentang Penetapan Daerah Tertinggal Tahun 2020-2024 (62 Daerah Tertinggal) bahwa layak untuk digunakan karena nilai *MSE* yang kecil yaitu 1% dan nilai *MAPE* yang berada di bawah 50%. Adapun hasil prediksi dari model prediksi ini adalah persentase penduduk miskin di daerah tertinggal dapat mencapai angka 35.03%. Dengan hasil prediksi tersebut, sangat diharapkan pemerintah Indonesia, khususnya daerah tertinggal dapat mengambil langkah tegas untuk menghentaskan kemiskinan. Selanjutnya, untuk melakukan prediksi yang lebih akurat dan presisi maka dibutuhkan jumlah data yang lebih besar. Karena pada dasarnya *machine learning* merupakan ilmu statistik yang diimplementasikan ke dalam bahasa pemrograman. Maka semakin banyak data yang dimasukkan, semakin baik mesin dapat membaca pola dari data tersebut.

UCAPAN TERIMA KASIH

Saya ucapkan terima kasih kepada seluruh dosen Teknik yang telah memberikan bantuan berupa pikiran dan cara menulis yang baik dan benar. Saya juga mengapresiasi kinerja dari BPS yang telah melakukan survey kepada Masyarakat dan membuka data tersebut untuk publik. Terakhir, saya ucapkan terima kasih kepada JIMSH LP3M yang telah meluangkan waktu untuk merevisi jurnal ini agar dapat laik untuk diterbitkan kepada publik.

DAFTAR PUSTAKA

- Alkharif, S., Lee, K., & Kim, H. (2018). *Time-Series Analysis for Price Prediction of Opportunistic Cloud Computing Resources*.
- Burkov, A. (2019). *The Hundred-Page Machine Learning Book*. Andriy Burkov.
- Choy, Y. T., Hoo, M. H., & Khor, K. C. (2021, September 8). Price Prediction Using Time-Series Algorithms for Stocks Listed on Bursa Malaysia. *2021 2nd International Conference on Artificial Intelligence and Data Sciences, AiDAS 2021*. <https://doi.org/10.1109/AiDAS53897.2021.9574445>
- Dwi Bekti, R. (2012). PREDIKSI DAN INTERPOLASI MELALUI ORDINARY KRIGING: STUDI KASUS KEMISKINAN DI PROVINSI JAWA TIMUR. *Jurnal Mat Stat*, 12(2), 123–132.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (R. Roumeliotis & N. Tache, Eds.; 2nd ed.). O'REILLY.
- Karim, A. (2020). Perbandingan Prediksi Kemiskinan di Indonesia Menggunakan Support Vector Machine (SVM) dengan Regresi Linear. *Jurnal Sains Matematika Dan Statistika*, 6(1).
- Kuhn, M., & Johnson, K. (2018). *Applied predictive modeling*. Springer.
- Mamase, S., Sinukun, R. S., Muchlis Rahim, J., Barat, P., Botupingge, K., & Bone Bolango, K. (2018). Prediksi Tingkat Kemiskinan Provinsi Gorontalo Menggunakan Metode Gabungan K-Means dan Generalized Regression Neural Network. *Edisi Nopember*, 8(2).
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MA: MIT Press.
- Raharjo, B. (2021). *Pembelajaran Mesin (Machine Learning)* (M. Wibowo Caroline, Ed.; 1st ed.). Yayasan Prima Agus Teknik.
- Rahimi, I., Gandomi, A. H., Asteris, P. G., & Chen, F. (2021). Analysis and prediction of covid-19 using sir, seiqr and machine learning models: Australia, italy and uk cases. *Information (Switzerland)*, 12(3), 1–25. <https://doi.org/10.3390/info12030109>
- Riady, S. R. (2023). Stock Price Prediction using Prophet Facebook Algorithm for BBKA and TLKM. *International Journal of Advances in Data and Information Systems*, 4(2). <https://doi.org/10.25008/ijadis.v4i2.1258>
- Syahrudin, Pujiana, E., Purnama Sari, I., Melia Mardika, V., & Putri, M. (2020). ANALISIS ALGORITMA BACK PROPAGATION DALAM PREDIKSI ANGKA KEMISKINAN DI INDONESIA. *Pendekar : Jurnal Pendidikan Berkarakter*, 3(1), 11–17. <https://doi.org/10.31764>
- Wanto, A. (2018). PENERAPAN JARINGAN SARAF TIRUAN DALAM MEMPREDIKSI JUMLAH KEMISKINAN PADA KABUPATEN/KOTA DI PROVINSI RIAU. *Kumpulan Jurnal Ilmu Komputer (KLIK)*, 05(01), 61–74.
- Wibawa, G. N. A., Yahya, I., Baharuddin, Rahman, G. A., & Agusrawati. (2022). ANALISIS REGRESI DATA PANEL PADA FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT KEMISKINAN SULAWESI TENGGARA TAHUN 2017-2020. *Jurnal Matematika, Komputasi Dan Statistika*, 2(3), 1–9.

Ye, Z. (2019). Air Pollutants Prediction in Shenzhen Based on ARIMA and Prophet Method. *E3S Web of Conferences*, 136. <https://doi.org/10.1051/e3sconf/20191360>