

## PREDIKSI JANGKA PANJANG COVID-19 INDONESIA MENGGUNAKAN DEEP LEARNING

### LONG-TERM PREDICTION FOR COVID-19 INDONESIA USING DEEP LEARNING

Adhitho Satyo Bayangkari Karno<sup>1</sup>, Widi Hastomo<sup>2</sup> dan Indra Sari Kusuma Wardhana<sup>3</sup>

Fakultas Teknologi Informasi, Universitas Gunadarma<sup>1</sup>, Fakultas Ekonomi dan Bisnis, ITB-Ahmad Dahlan Jakarta<sup>2</sup>, Magister Teknologi Informasi STMIK Jakarta STI&K<sup>2,3</sup>  
Adh1t10.2@gmail.com

**Abstract:** Lack in machine learning methods are the inability to predict more than one step forward, because in each step forward machine learning must calculate the error value with real data to be used to correct the weight value in the next round. This study aims to predict the long-term future of Covid-19 in Indonesia. The method used is to make the results of one-step predictions from machine learning as the first new data. Then in the next iteration machine learning will produce one more prediction step as the second new data, so we get two prediction results going forward. This process is repeated until it reaches the desired long-term prediction of 2 months (50 days). Long prediction for the next 2 months, is done by using 2 Deep Learning (DL) methods, namely Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) in the 4 hidden layers of learning machine used. The GRU model on the four hidden layers gives the best results with a value of RMSE = 206,632 at epoch = 5.

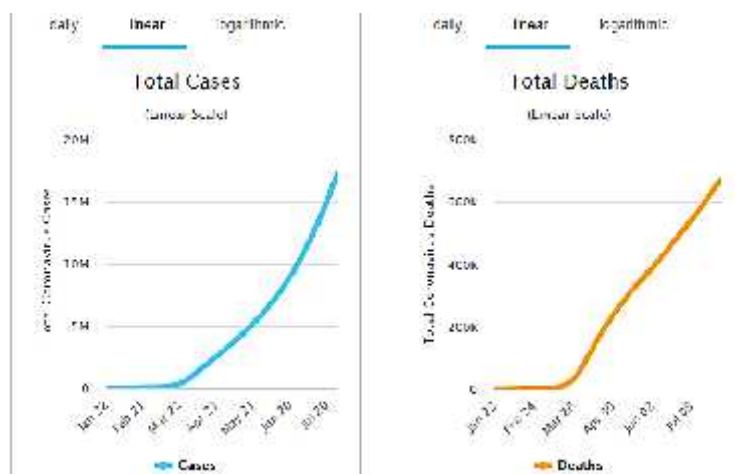
**Keywords:** Covid-19, Deep Learning, GRU, LSTM

**Abstrak:** Kelemahan dalam metode mesin learning adalah ketidakmampuan dalam memprediksi lebih dari satu langkah kedepan, dikarenakan dalam setiap langkah kedepan mesin learning harus menghitung nilai error dengan data real untuk dipergunakan memperbaiki nilai bobot dalam putaran selanjutnya. Penelitian ini bertujuan untuk memprediksi jangka panjang kedepan dari Covid-19 di Indonesia. Metode yang dipergunakan adalah menjadikan hasil prediksi satu langkah dari mesin learning sebagai data baru pertama. Kemudian dalam iterasi berikutnya mesin learning akan menghasilkan satu langkah prediksi lagi sebagai data baru ke dua, sehingga didapat dua hasil prediksi kedepan. Proses ini diulangi terus hingga mencapai prediksi jangka panjang yang diinginkan yaitu 2 bulan kedepan (50 hari). Prediksi panjang untuk 2 bulan kedepan, dilakukan dengan menggunakan 2 metode Deep Learning (DL) yaitu Long Short Term Memory (LSTM) dan Gated Recurrent Unit (GRU) di 4 hidden layer mesin learning yang digunakan. Model GRU pada ke empat lapisan hidden layer memberikan hasil terbaik dengan nilai RMSE = 206.632 di epoch = 5.

**Kata kunci:** Covid-19, Deep Learning, GRU, LSTM

### PENDAHULUAN

Pertama kali covid-19 dilaporkan oleh pemerintah China pada akhir desember 2019, sampai hari ini tanggal 31 Juli 2020 sudah menyebar di 213 negara didunia, dengan total kasus mencapai 17.345.261 juta jiwa (Anonim, 2020). Sedangkan kasus covid-19 pertama kali diumumkan oleh Presiden Jokowi pada awal maret 2020, hingga hari ini jumlah kasus di Indonesia mencapai 106.336 ribu jiwa, dalam kurun waktu 5 bulan, penyebaran sangat cepat, rerata perhari kasus yang terinfeksi mencapai 709 kasus. Pandemi Covid-19 telah banyak merubah kehidupan manusia di seluruh dunia dalam berbagai sendi-sendi kehidupannya. Pandemi ini telah membawa manusia ke dalam kehidupan yang tidak normal dari masa-masa sebelumnya, dibutuhkan adaptasi agar mampu bertahan pada situasi seperti sekarang ini. Belum ditemukannya vaksin covid-19 memaksa setiap manusia hidup berdampingan dengan virus tersebut. Pada grafik 1 terlihat tingginya penyebaran kasus baru dan angka kematian di seluruh dunia



Grafik 1. Kasus baru dan kematian di dunia (Anonim, 2020)

Dampak dari pandemi covid-19 di Indonesia terasa sangat dalam, hal ini terlihat dari daya beli masyarakat menurun signifikan, banyak perusahaan yang gulung tikar, menambah angka pengangguran, saat ini memasuki masa resesi dan diprediksi masa resesi ini lebih berat dibandingkan dengan krisis 1998. Pada kuartal kedua pemerintah memproyeksi ekonomi Indonesia akan minus hingga 3,8%. Banyak pihak sudah mencoba untuk memprediksi kapan berakhirnya pandemi covid-19, bagaimana memutus rantai penyebaran covid-19 dengan metode matematika (Shah, Abdeljawad, Mahariq, Jarad, & Deniz, 2020),(Zeb, Alzahrani, Erturk, & Zaman, 2020), statistika (Zuo, Khosa, Ahmad, Almaspoor, & Jo, 2020) hingga mencari vaksin penangkalnya, namun banyak tantangan yang dihadapi agar prediksi tersebut menjadi valid, antara lain yaitu kepatuhan terhadap protokol covid-19.

Pada penelitian ini ingin mencoba prediksi penyebaran kasus baru covid-19, agar dapat mengetahui kondisi pada beberapa hari kedepan, dengan kendala ketersediaan data yang minim menjadi tantangan utama dalam penelitian menggunakan mesin learning ini. Dengan melakukan berbagai kombinasi dan variasi dari data dan metode Deep Learning yang ada, diharapkan dapat memberikan prediksi dengan akurasi sebaik mungkin. Dasar metode Deep Learning (DL) adalah neural network, berupa rangkaian yang terdiri dari 1 input layer, beberapa hidden layer dan 1 output layer. Setiap hidden layer terdiri dari beberapa node, dimana setiap node akan menerima semua input dari layer sebelumnya dengan kelipatan bobot ditambah nilai bias sebagai nilai awal. Hasil (output) setiap node akan menjadi input kembali ke node di layer selanjutnya, dan akhir dari proses ini adalah nilai prediksi yang akan dibandingkan dengan nilai target (nilai real). Perbandingan nilai prediksi dengan nilai target berupa nilai kesalahan (error) yang dijadikan sebagai acuan untuk memperbaiki (update) nilai bobot baru yang akan dipergunakan untuk putaran (iterasi) selanjutnya. Putaran ini dilakukan terus menerus hingga mencapai nilai error yang minimum.

Prediksi DL menggunakan data time series akan menghasilkan hanya satu langkah prediksi (pertama). Untuk prediksi langkah selanjutnya (kedua) tidak dapat dilakukan, karena tidak adanya nilai real yang dipergunakan untuk menghitung nilai error dan bobot untuk putaran berikutnya. Hal ini menjadi kekurangan dalam proses prediksi menggunakan DL, untuk memperoleh prediksi jangka panjang lebih dari sekedar prediksi jangka pendek satu langkah ke depan. Untuk mengatasi kekurangan tersebut serta memperoleh prediksi jangka panjang dari data time series Covid-19, maka hasil prediksi pertama dipergunakan sebagai data baru. Dengan adanya data baru ini prediksi kedua dapat dilakukan, hasil dari prediksi kedua dijadikan sebagai data baru ke dua. Proses ini dilakukan berulang hingga mencapai prediksi jangka panjang yang diinginkan dalam hal ini adalah sekitar 2 bulan (50 hari) kedepan.

Kebaruan dari penelitian ini yaitu; (a) mampu memprediksi penyebaran Covid-19 hingga 2 bulan kedepan, (b) mengkombinasikan hidden layer dengan menggunakan variasi kedua model LSTM dan GRU, (c) mengetahui karakteristik epoch-rmse untuk setiap dataset,

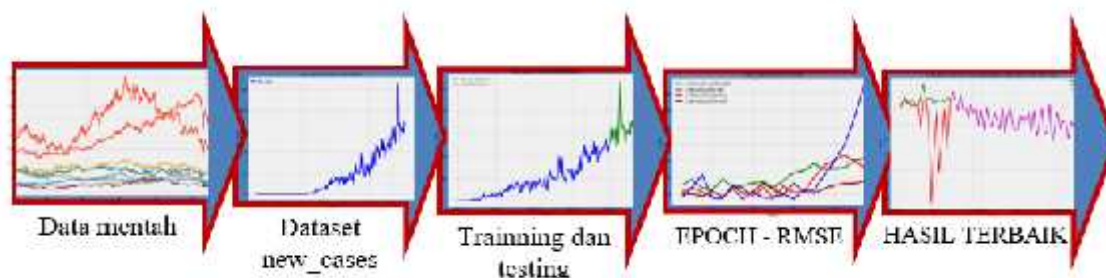
sehingga dapat diperoleh nilai epoch yang tepat untuk menghasilkan prediksi yang lebih akurat.

## PENELITIAN TERKAIT

(Islam, Islam, & Asraf, 2020) berpendapat bahwa hasil prediksi image x-ray covid-19 dengan kombinasi CNN dengan LSTM menghasilkan akurasi mencapai 97%, spesifisitas 91%, dan sensitivitas 100% namun kendala pada keterbatasan jumlah data image. Dalam penelitian (Car, S, An, Lorencin, & Mrzljak, 2020) prediksi penyebaran covid-19 menggunakan dataset dari 406 lokasi dalam waktu 51 hari, metode yang digunakan multilayer perceptron (MLP) artificial neural network (ANN) mendapatkan hasil  $R^2$  98,5% kasus terkonfirmasi, 99,4% kasus meninggal dan 98% pasien yang pulih. Hasil dari penelitian (Car et al., 2020) prediksi penyebaran covid-19 menggunakan Support Vector Regression and stacking-ensemble menggunakan data clinical, jumlah objek 40.581 menghasilkan akurasi eror mencapai 0.87%-3.51% satu hari, 1.02%-5.63% tiga hari dan 0.95% -6.90% enam hari kedepan. Sedangkan penelitian (Chimmula & Zhang, 2020) menggunakan metode LSTM, prediksi covid-19 mencapai hari ke 14 untuk eror mortality rate 3,2% sedangkan recovery rate mencapai 2,5%.

## METODE

Dalam penelitian ini menggunakan 4 model arsitektur hidden layer yang berbeda, dengan melakukan variasi LSTM dan GRU. Hal ini dimaksudkan untuk memperoleh arsitektur hidden layer dengan hasil akurasi yang lebih baik (RMSE minimum). Selain itu tindakan ini juga memudahkan dalam menentukan besaran nilai epoch yang diperlukan sebelum proses prediksi dengan data sebenarnya dilakukan. Sehingga tidak perlu lagi melakukan coba salah (try-error) atas penggunaan nilai epoch yang diperlukan untuk proses prediksi dengan DL. Semua bahasa program yang dipergunakan dalam penelitian ini menggunakan bahasa Python. Bahasa ini dipilih karena banyak kemudahan yang dapat dilakukan, terutama untuk jenis pemrograman yang berhubungan dengan mesin learning. Library yang ada di Python memudahkan dalam melakukan manipulasi vektor dan matrik (seperti : numpy(Dabhi & Parmar, 2020; Faouzi & Janati, 2020; Grout, 2020; Santos & Brussel, 2019), pandas (Bantilan, 2020; Ford, Weitzner, & Bahl, 2020), tensor flow (Andrade-Loarca & Kutyniok, 2020; Grattarola & Alippi, 2020; Lagouvardos, Dolby, & Grech, 2020), keras (Grattarola & Alippi, 2020; Ott et al., 2020; Yanwei HYang, Xiaojian Hao, 2020)), output dapat ditampilkan dalam berbagai macam bentuk yang menarik baik secara visual ataupun numerik (matplotlib (Adhitho Satyo Bayangkari Karno, 2020; Nielsen, 2013; Pawase & Raja, 2020), seaborn (Flores & Flores, 2020; Lock, Bednarz, Pettit, & Lock, 2020; Mohanty, Mishra, & Saxena, 2020; Nielsen, 2013), fbprophet (Hastomo & Satyo, 2019)), dan penghitungan kesalahan (MSE, RMSE) dengan menggunakan library sklearn. Pada gambar 1 skenario penelitian kombinasi antara DL Lstm dan Gru.



Gambar 1. Skenario Penelitian

Data mentah (sumber "ourworldindata.org") adalah data seluruh manca negara sebagai berikut;

Tabel 1. Data mentah

iso_code	continent	location	total_cases	new_cases	total_deaths	new_deaths	total_cases_per_million	new_cases_per_million	total_recovered
2019-12-31	AFR	Adic. Afghanistan	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2020-01-01	AFR	Adic. Afghanistan	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2020-01-02	AFR	Adic. Afghanistan	0.0	0.0	0.0	0.0	0.0	0.0	0.0

3 rows • 33 columns

(2147, 11)

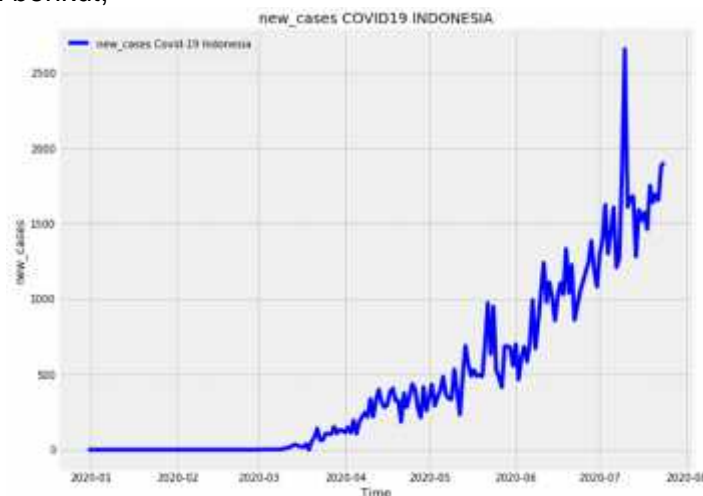
Data mentah (31 Desember 2019 - 24 Juli 2020) berisi 33 fitur kolom dan 32,374 baris, dilakukan filter untuk memperoleh data hanya negara Indonesia dan mengambil fitur kolom “new\_cases”.

**Tabel 2.** Data new\_cases Covid-19 Indonesia

new_cases		new_cases	
date	new_cases	date	new_cases
2019-12-31	0.0	2020-07-22	1655.0
2020-01-01	0.0	2020-07-23	1682.0
2020-01-02	0.0	2020-07-24	1606.0

(200, 33)

Hasil filter adalah data Indonesia dengan fitur new\_cases berjumlah 200 baris data, dengan grafik sebagai berikut,



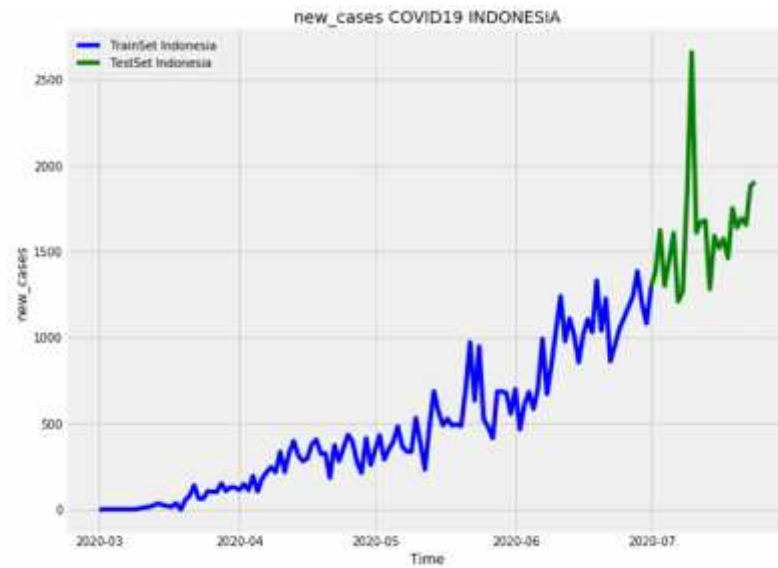
**Grafik 2.** Grafik new\_cases Covid-19 Indonesia

Untuk menyediakan data ke mesin agar dapat melakukan pelatihan dan pengujian hasil pembelajaran, maka data dibagi menjadi training dataset sebanyak 116 dan testing dataset sebanyak 24, mulai dari tanggal 1 Maret 2020 sampai dengan 24 Juli 2020.

**Tabel 3.** Data Training dan testing dataset

new_cases		new_cases		new_cases		new_cases	
date	new_cases	date	new_cases	date	new_cases	date	new_cases
2020-03-01	0.0	2020-06-29	1148.0	2020-07-01	1293.0	2020-07-22	1655.0
2020-03-02	0.0	2020-06-30	1052.0	2020-07-02	1355.0	2020-07-23	1682.0
2020-03-03	2.0	2020-07-01	1253.0	2020-07-03	1521.0	2020-07-24	1606.0

(116, 1) (24, 1)

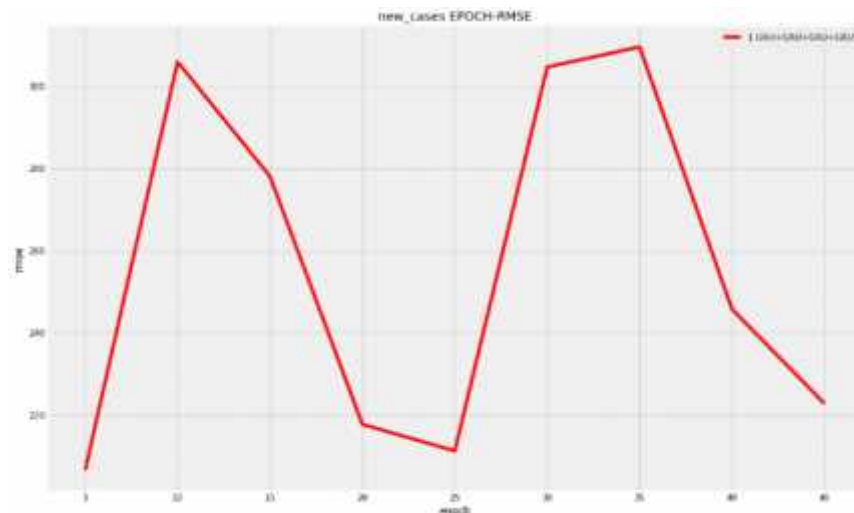


**Grafik 3.** Dataset Training dan testing

Salah satu tantangan untuk memperoleh hasil prediksi yang akurat adalah menentukan nilai parameter epoch yang tepat. Untuk menghindari tindakan coba-salah (try-error) dengan hasil tidak menentu, maka sebelum melakukan proses DL dibuat suatu program iterasi untuk mengetahui nilai epoch dengan RMSE yang minimum menggunakan 2 model hidden layer yang berbeda. Model hidden layer pertama adalah 4 hidden layer berisi LSTM dan model hidden layer kedua berisi 4 hidden layer GRU. Hasil dari program iterasi ini adalah grafik yang ditunjukkan dibawah ini,



**Grafik 4.** Epoch vs RMSE untuk model 0

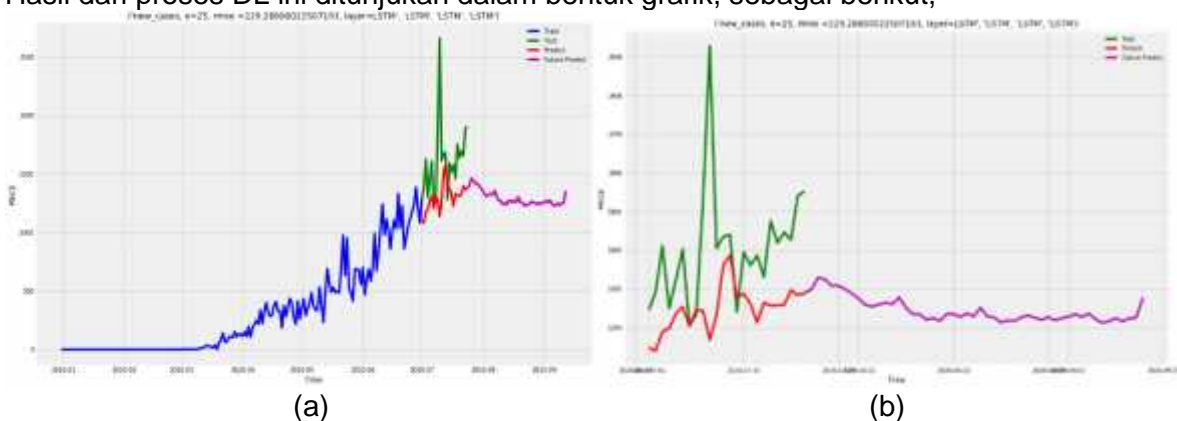


Grafik 5. Epoch vs RMSE untuk model 1

Terlihat dari Grafik 5, epoch yang tepat untuk model 0 (4 hidden layer LSTM) adalah 25 dengan nilai RMSE = 229.288. Grafik 5, epoch yang tepat untuk model 1 (4 hidden layer GRU) adalah 5 dengan nilai RMSE = 206.632.

### HASIL EKSPERIMEN

Dari Grafik karakteristik Epoch vs RMSE, kemudian mulai dilakukan proses DL menggunakan epoch dengan nilai RMSE minimum berdasarkan grafik (Grafik 6 dan 7). Hasil dari proses DL ini ditunjukkan dalam bentuk grafik, sebagai berikut,



Grafik 6. (a) prediksi model 0 (keseluruhan), (b) prediksi model 0 (diperbesar)



Grafik 7. (a) Prediksi model 1 (keseluruhan), (b) Prediksi model 1 (diperbesar)

## SIMPULAN

1. Dari 2 model arsitektur hidden layer terlihat bahwa model 1 dengan 4 hidden layer berisi GRU dapat menghasilkan prediksi yang lebih baik dibandingkan model 0 (4 hidden layer LSTM).
2. Kekurangan karena minimnya data dapat diatasi dengan memperkecil nilai time step, sehingga data menjadi lebih banyak sehingga memungkinkan proses DL dapat menjadi lebih baik.
3. Untuk selanjutnya penelitian ini dapat dikembangkan dengan variasi model yang berbeda untuk memperoleh kemungkinan nilai RMSE dan epoch yang lebih kecil lagi, sehingga prediksi akan menjadi lebih akurat.
4. Untuk dapat mengetahui berakhirnya masa pandemi, perlu dilakukan penelitian kembali untuk prediksi kedepan yang lebih jauh lagi.

## DAFTAR PUSTAKA

- Adhitho Satyo Bayangkari Karno. (2020). *Prediksi Data Time Series Saham Bank BRI Dengan Mesin Belajar LSTM ( Long ShortTerm Memory )*. 1(1), 1–8.
- Andrade-Loarca, H., & Kutyniok, G. (2020). *tfShearlab: The TensorFlow Digital Shearlet Transform for Deep Learning*. 1–25. Retrieved from <http://arxiv.org/abs/2006.04591>
- Anonim. (2020). COVID-19 Coronavirus Pandemic. Retrieved July 31, 2020, from Worldometers website: <https://www.worldometers.info/coronavirus/>
- Bantilan, N. (2020). *pandera: Statistical Data Validation of Pandas Dataframes*. (Scipy), 106–114.
- Car, Z., Š, S. B., An, N., Lorencin, I., & Mrzljak, V. (2020). *Modeling the Spread of COVID-19 Infection Using a Multilayer Perceptron*. 2020.
- Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons and Fractals*, 135. <https://doi.org/10.1016/j.chaos.2020.109864>
- Dabhi, S., & Parmar, M. (2020). *Eigenvector Component Calculation Speedup over NumPy for High-Performance Computing*. Retrieved from <http://arxiv.org/abs/2002.04989>
- Faouzi, J., & Janati, H. (2020). Pyts: A python package for time series classification. *Journal of Machine Learning Research*, 21, 1–6.
- Flores, A., & Flores, H. (2020). *Sentiment Analysis on Airline Tweets Using Naïve Bayes Classifier*.
- Ford, A. S., Weitzner, B. D., & Bahl, C. D. (2020). Integration of the Rosetta suite with the python software stack via reproducible packaging and core programming interfaces for distributed simulation. *Protein Science*, 29(1), 43–51. <https://doi.org/10.1002/pro.3721>
- Grattarola, D., & Alippi, C. (2020). *Graph Neural Networks in TensorFlow and Keras with Spektral*. Retrieved from <http://arxiv.org/abs/2006.12138>
- Grout, I. (2020). Realization of NumPy TensorDot using the Field Programmable Gate Array for Embedded Machine Learning Applications. *2020 8th International Electrical Engineering Congress, IEECON 2020*. <https://doi.org/10.1109/IEECON48109.2020.229523>
- Hastomo, W., & Satyo, A. (2019). *Long Short Term Memory Machine Learning Untuk Memprediksi Akurasi Nilai Tukar IDR Terhadap USD*. 3.
- Islam, M. Z., Islam, M. M., & Asraf, A. (2020). *A Combined Deep CNN-LSTM Network for the Detection of Novel Coronavirus ( COVID-19 ) Using X-ray Images*. (June), 1–20. <https://doi.org/10.1101/2020.06.18.20134718>
- Lagouvardos, S., Dolby, J., & Grech, N. (2020). Static Analysis of Shape in TensorFlow Programs. *34th European Conference on Object-Oriented Programming (ECOOP 2020)*, (15), 1–15.
- Lock, O., Bednarz, T., Pettit, C., & Lock, O. (2020). The visual analytics of big , open public transport data – a framework and pipeline for monitoring system performance in Greater Sydney. *Big Earth Data*, 00(00), 1–26. <https://doi.org/10.1080/20964471.2020.1758537>

- Mohanty, S., Mishra, A., & Saxena, A. (2020). *EasyChair Preprint Medical Data Analysis Using Machine Learning with KNN MEDICAL DATA ANALYSIS USING MACHINE*.
- Nielsen, A. (2013). *Python programming — plotting*. IX(li), 26.
- Ott, J., Pritchard, M., Best, N., Linstead, E., Curcic, M., & Baldi, P. (2020). *A Fortran-Keras Deep Learning Bridge for Scientific Computing*. 1–14. Retrieved from <http://arxiv.org/abs/2004.10652>
- Pawase, R., & Raja, K. S. (2020). *pyCallisto: A Python Library To Process The CALLISTO Spectrometer Data*. 1–14. Retrieved from <http://arxiv.org/abs/2006.16300>
- Santos, R., & Brussel, V. U. (2019). *Awkward Array: JSON-like data, NumPy-like idioms*. (October 2016), 49–56.
- Shah, K., Abdeljawad, T., Mahariq, I., Jarad, F., & Deniz, S. (2020). Qualitative Analysis of a Mathematical Model in the Time of COVID-19. *BioMed Research International*, 2020(December 2019). <https://doi.org/10.1155/2020/5098598>
- Yanwei HYang, Xiaojian Hao, L. Z. and L. R. (2020). Application of Scikit and Keras Libraries for the Classification of Iron Ore Data Acquired by Laser-Induced Breakdown Spectroscopy (LIBS). *Mdpi Journal Sensors*, 1393.
- Zeb, A., Alzahrani, E., Erturk, V. S., & Zaman, G. (2020). *Mathematical Model for Coronavirus Disease 2019 ( COVID-19 ) Containing Isolation Class. 2020*.
- Zuo, M., Khosa, S. K., Ahmad, Z., Almaspoor, Z., & Jo, J. (2020). *Comparison of COVID-19 Pandemic Dynamics in Asian Countries with Statistical Modeling. 2020*.