

PERBANDINGAN PERFORMA MODEL MACHINE LEARNING DALAM PREDIKSI SUHU DI SEMARANG

Apit Sutaryani^{1*}, Sunarno Sunarno², Djuniadi Djuniadi³

^{1,2} Fisika Universitas Negeri Semarang

³ Teknik Elektro Universitas Negeri Semarang, Indonesia

Received: 11 Juli 2024

Accepted: 31 Juli 2024

Published: 7 Agustus 2024

Keywords:

3-5 keyword;

prediksi, machine learning, regresi linier, decision tree

Correspondent Email:

¹pitsutar@students.unnes.ac.id

²NarnoPhysics@mail.unnes.ac.id

³ djuniadi@mail.unnes.ac.id

Abstrak. Prediksi suhu udara sangat bermanfaat untuk pertimbangan kebijakan lokal, seperti suhu udara lebih tinggi dari tahun-tahun sebelumnya, sehingga perlunya perluasan lahan hijau sebagai strategi untuk memperbaiki kualitas udara dan meningkatkan kualitas hidup masyarakat. Oleh karena itu perencanaan dan implementasi proyek lahan hijau harus menjadi prioritas dalam upaya pengembangan kota yang berkelanjutan. Penelitian ini membandingkan algoritma *machine learning* model *regresi linier* dengan *decision tree* untuk memprediksi suhu kota Semarang dari tahun 2019 hingga 2023. Variabel data yang digunakan meliputi temperatur, titik embun, kecepatan angin, curah hujan, tekanan udara, lamanya penyinaran matahari, dan kelembaban. Pengolahan data dilakukan menggunakan software *rapid miner* dengan menggunakan algoritma *regresi linier* dan *decision tree*. Hasil penelitian menunjukkan algoritma regresi linier memiliki nilai RMSE sebesar 0.131 +/- 0.000, MAE 0,099 +/- 0,086 dan R2 0.990, sedangkan pada algoritma *decision tree* nilai RMSE sebesar 0,293 +/- 0,000, nilai MEA 0,189 +/- 0,224 dan nilai R2 0,948. Kesimpulan analisis menunjukkan bahwa algoritma regresi linier lebih akurat daripada *decision tree*.

Abstract. *Forecasting the temperature of the air is very useful for local policy considerations, it seems that the air temperature is higher than in previous years, so the need for greenery expansion as a strategy to improve air quality and improve the quality of life of people. Therefore, planning and implementation of green land projects should be a priority in sustainable urban development efforts. The study compared the algorithm machine learning linear regression model with the decision tree to predict the temperature of Semarang city from 2019 to 2023. The data variables used include temperature, dew point, wind intensity, rainfall, air pressure, sunlight duration, and humidity. Data processing is done using rapid miner software using linear regression algorithms and decision trees. The results of the research showed that the linear regression algorithm has RMSE values of 0.131 +/- 0.000, MAE of 0.099 +/- 0.086 and R2 of 0.990, whereas on the decision tree algorithm the RMSE value is 0.293 +/- 0.000, the MEA value of 0.189 +/- 0.224 and the R2 value of 0.948. Analysis conclusions show that linear regression algorithms are more accurate than the decision tree.*

1. PENDAHULUAN

Prediksi (peramalan) adalah upaya untuk mengetahui apa yang akan terjadi di masa depan dengan melihat apa yang telah terjadi sebelumnya. Ini dilakukan dengan mengumpulkan sejarah data, atau jumlah data sebelumnya, dan kemudian melakukan analisis dilanjutkan dengan penghitungan data tersebut untuk menghasilkan data prediksi untuk masa depan. [1].

Prediksi suhu merupakan aspek krusial dalam berbagai bidang, termasuk pertanian, perencanaan kota, manajemen energi, dan mitigasi bencana [2]. Sebagai ibu kota Provinsi Jawa Tengah, Indonesia, Semarang memiliki iklim tropis dengan karakteristik cuaca yang beragam sepanjang tahun [3]. Wilayah Kota Semarang dikatakan mempunyai karakteristik geografis yang unik dengan pesisir di utara dan pegunungan di selatan serta memiliki iklim tropis basah [4][5]. Disamping itu Wilayah

Kota Semarang juga mempunyai topografi yang bervariasi dan curah hujan yang tinggi.[6].

Perubahan suhu yang tak terduga dapat mempengaruhi kehidupan sehari-hari masyarakatnya, aktivitas ekonomi, serta berbagai sektor lainnya. Oleh karena itu, prediksi suhu yang akurat sangat dibutuhkan untuk meningkatkan pengambilan keputusan dan perencanaan.[7].

Suhu adalah komponen penting dari iklim dan cuaca, dan sangat memengaruhi kehidupan manusia dan lingkungan. Masa depan iklim akan ditunjukkan oleh prediksi suhu. Dibutuhkan alat untuk menganalisa dan memprediksi cuaca. *Machine learning* adalah teknologi yang tepat untuk digunakan dalam situasi ini. [8]

Teknologi machine learning telah membuka peluang baru dalam meningkatkan akurasi prediksi suhu. Model-model *machine learning* menawarkan pendekatan yang lebih fleksibel dan adaptif dibandingkan metode statistik tradisional.

Machine learning adalah bidang ilmu yang mencakup perancangan dan pengembangan algoritma yang memungkinkan komputer untuk mengubah perilaku berdasarkan data yang diberikan. Metode untuk membuat model algoritma ini dikenal sebagai pembelajaran (learning) otomatis mesin. [9]

Model *regresi linier* dan *decision tree* memungkinkan sistem komputer untuk mempelajari pola-pola dari data dan menghasilkan prediksi yang akurat. Berkaitan dengan konteks prediksi cuaca dan iklim, *machine learning* dapat membantu mengidentifikasi pola-pola yang berulang dan menghasilkan prediksi yang lebih akurat berdasarkan data historis.[10]

Prediksi suhu menggunakan *machine learning* maka kualitas dan kuantitas data yang digunakan memiliki peran yang sangat penting. [11] Data historis cuaca dan iklim menjadi komponen utama yang digunakan sebagai input untuk sistem *machine learning*. Data tersebut mencakup berbagai parameter meliputi suhu udara, titik embun, kelembaban, curah hujan, tekanan udara, penyinaran matahari dan kecepatan angin.

Terdapat banyak algoritma *machine learning* ataupun *deep learning* untuk prediksi cuaca dan iklim maupun salah satu unsur atau variabel [12][13][14], *deep learning* selain

dapat memprediksi cuaca atau parameternya dapat juga diimplementasikan untuk perkiraan energi terbarukan atau dapat menganalisa *object unstructure* [15][16] Penelitian ini menggunakan 2 metode *machine learning* yaitu *Regresi Linear* dan *Decision Tree*. Penelitian ini bertujuan untuk melakukan perbandingan performa dua model machine learning dalam memprediksi suhu di Semarang. Studi perbandingan dua model tersebut, diharapkan dapat ditemukan model yang paling efektif dan akurat dalam konteks geografis dan iklim di Semarang.

2. TINJAUAN PUSTAKA

Tempat penelitian yang dipilih adalah Kota Semarang dengan koordinat 110.4222301937964 BT-6.989896609324934 LS.

Data yang digunakan merupakan data harian dari Stasiun Meteorologi Ahmad Yani Semarang selama 5 tahun dari 1 Januari 2019 - 31 Desember 2023.

2.1 Data Penelitian

Data-data yang digunakan untuk prediksi suhu di Semarang meliputi,

- Data Temperatur adalah data Suhu, data disini data yang diperoleh dalam interval waktu per hari. (Tricahyono, 2007).
- Data DEW (*Dew point/* titik embun) adalah suhu di mana udara jenuh dengan uap air, sehingga udara tidak dapat menahan lebih banyak uap air tanpa kondensasi. Titik embun merupakan indikasi kelembaban udara, dimana semakin tinggi titik embun, semakin tinggi kelembaban udara. Titik embun digunakan dalam model peramalan cuaca untuk memprediksi kondisi kelembaban udara di masa depan.
- Data kecepatan angin adalah kecepatan pergerakan udara horizontal di permukaan bumi. Prediksi kecepatan dan arah angin membantu dalam meramalkan kondisi cuaca di masa depan.
- Jumlah air yang jatuh dari atmosfer ke permukaan bumi dalam bentuk hujan, salju, hujan es, atau bentuk presipitasi

lainnya dalam jangka waktu tertentu dikenal sebagai data curah hujan. Data curah hujan digunakan dalam peramalan cuaca untuk memprediksi jumlah dan intensitas hujan di masa depan.

- e. Data PS (*Pressure*/Tekanan Udara) adalah kekuatan yang dihasilkan oleh berat udara di atas suatu titik dalam atmosfer, dan biasanya diukur dalam satuan seperti *hectopascal* (hPa) atau *millibar* (mb). Perubahan tekanan udara dapat menunjukkan perubahan cuaca yang akan datang.
- f. Data MAT (Matahari) adalah data yang mengukur durasi dan intensitas cahaya matahari yang diterima di suatu lokasi selama periode waktu tertentu.
- g. Data Kelembaban adalah kekuatan yang dihasilkan oleh berat udara di atas suatu titik dalam atmosfer, dan biasanya diukur dalam satuan seperti *hectopascal* (hPa) atau *millibar* (mb). Perubahan tekanan udara dapat menunjukkan perubahan cuaca yang akan datang.

2.2 Model Machine learning

Tujuan *machine learning* dalam kecerdasan buatan adalah untuk memahami bagaimana data terstruktur dan mengubahnya menjadi model. *Supervised learning* adalah bagian dari *machine learning* yang bertujuan untuk menggunakan data yang ada untuk membuat fungsi input-output atau hubungan input dan output. (Tagliaferri, 2017).

Algoritma yang belajar dari data latihan yang sudah dilabeli juga disebut *supervised learning*. dengan tujuan melakukan generalisasi data input (Iguar & Segui, 2017). *Supervised learning* bertugas membangun fungsi $y=f(x)$ dimana y adalah output dan x adalah input dengan data x bisa lebih dari 1 parameter. Data yang disediakan untuk membangun fungsi disebut data latihan, data ini digunakan untuk memprediksi data output. Sedangkan data yang digunakan untuk menguji akurasi sistem atau fungsi disebut dengan data uji.

Salah satu bagian dari *supervised learning* yang paling dikenal adalah model

regresi linier. Pada penelitian ini, model *machine learning* yang digunakan, yaitu :

1. *Regresi linier*. Nilai data yang terkait dan diketahui digunakan dalam *regresi linier* dalam analisis data untuk memprediksi nilai data yang tidak diketahui. (Suspeno, 2009). *Regresi linier* merupakan hubungan linear (dalam parameter) antara *dependent variable* (y_i) dan *explanatory variables* (x_i) = (x_{i0} , x_{i1} , ..., x_{iK}).
 x_{i0} : 1 (konstan kecuali dinyatakan lain
 x_{ik} : *independent variable*
 i : 1, ..., N_i

$$Y_t = X_t \beta + \epsilon_t \quad (1)$$

Y_t : nilai estimasi pada waktu t (variabel respon)

β : (β_0 , β_1 , ..., β_k) indikator dari hubungan antara Y_t and X_t ,

X_t : (1 , x_{1t} , x_{2t} , . . . , x_{jt}) variabel prediktor

ϵ_t : istilah kesalahan acak pada waktu t dimana $t = 1, \dots, N$.

2. *Decision Tree* (Perintah Keputusan).

Metode pohon keputusan adalah bagian dari pendekatan fungsi target bernilai diskrit. Ini paling umum digunakan dalam pembelajaran mesin dasar (Grosan & Abraham, 2011).

Formula berikut digunakan untuk menentukan akar berdasarkan nilai gain tertinggi dari atribut yang tersedia. Sebelum mendapatkan nilai gain, nilai entropy harus ditentukan. Rumus Entropy :

$$Entropy(s) = \sum_{i=1}^n -p_i \log_2 p_i \quad (2)$$

Dimana

S : Himpunan Kasus

n : Jumlah partisi S

P_i : Proporsi dari S_i terhadap S

Rumus Gain :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S_i) \quad (3)$$

Dimana
 S : Himpunan Kasus
 A : Atribut
 N : Jumlah Partisi Atribut a
 |S| : Jumlah Kasus pada Partisi ke-1
 |S| : Jumlah Kasus dalam S

3. METODE PENELITIAN

Penelitian ini menggunakan teknik *machine learning* dengan algoritma *Regresi Linier* dan *Decision tree* dengan tujuan membandingkan dua algoritma tersebut sehingga diperoleh nilai prediksi yang lebih akurat.

Pengolah data untuk menghasilkan prediksi pada penelitian ini menggunakan software rapid miner. Software tersebut salah satu untuk menerapkan metode *machine learning* dalam suatu prediksi dan untuk mengevaluasi hasil prediksi dapat menggunakan matriks statistik RMSE (*Root Mean Squared Error*) dan MAE (*Mean Absolute Error*) serta R² (*Squared Correlation*).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (4)$$

Dimana ;
 y_i = Nilai data pengamatan
 y'_i = Nilai data prakiraan
 n = Jumlah data

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x'_i| \quad (5)$$

Dimana ;
 x_i = Nilai data sebenarnya
 x'_i = Nilai data prediksi
 n = Jumlah data

$$R^2 = 1 - \frac{\sum_{i=1}^n (z_i - z'_i)^2}{\sum_{i=1}^n (z_i - \hat{z})^2} \quad (6)$$

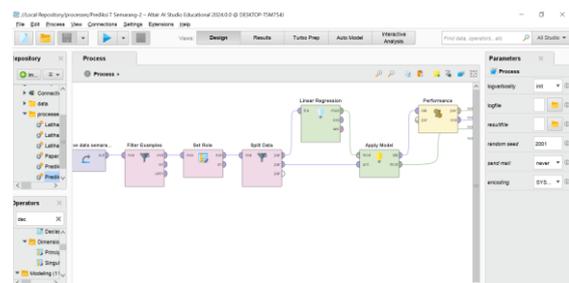
Dimana ;
 z_i = Data aktual
 z'_i = Data prediksi
 \hat{z} = Rata – rata dari data aktual
 n = Jumlah data

4. HASIL DAN PEMBAHASAN

Data penelitian yang terdiri dari variabel temperatur, titik embun, kecepatan angin, curah hujan, tekanan udara, lamanya penyinaran matahari dan kelembaban selama periode tahun 2019 – 2023 kemudian data tersebut di import ke dalam software rapid miner untuk dilakukan pengolahan.

Setelah data dinyatakan lengkap maka langkah berikutnya menentukan parameter yang akan di prediksi menggunakan dua algoritma yaitu regresi linier dan decision tree dengan komposisi data uji 30% dan data latihan 70%.

Tahapan pengolahan dengan algoritma regresi linier terlihat pada gambar 1 dan hasil prediksi temperatur dapat dilihat di gambar 2 sedangkan performa dari algoritmanya lihat tabel 1



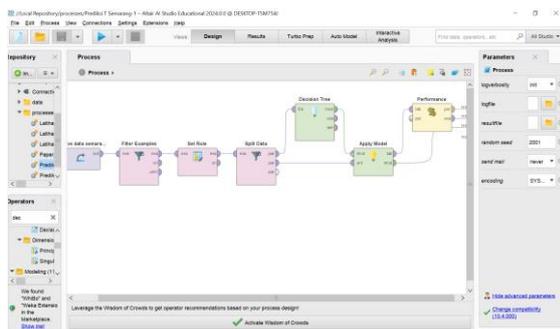
Gambar 1 Pengolahan dengan Regresi Linier

Gambar 2 Hasil Prediksi dengan Regresi Linier

Tabel 1 Performace Regresi Linier

No.	Uji Performance	Nilai
1.	Root Mean Squared Error (RMSE)	0.131 +/- 0.000
2.	Mean Absolute Error (MAE)	0.099 +/- 0.086
3.	Squared Correlation (R ²)	0.990

Gambar 3 menunjukkan tahapan pengolahan dan hasil prediksi temperatur dengan algoritma decision tree, dan tabel 2 menunjukkan kinerja algoritma.



Gambar 3 Pengolahan dengan Decision Tree

Row No.	Temp	gradation/Tem.	Wind Speed	Dew Point	Rainfall	Pressure G.	Relative H.	Burntime H.
1	28.800	28.400	2.100	24.900	0	1008.100	81.200	2.800
2	29	29.100	1.900	24.200	0	1010.500	79.400	0
3	29.300	29.233	1.900	25.300	0	1011.400	79.100	0
4	30.600	30.700	1.700	25.300	0	1010.900	73.900	9
5	29.700	29.720	2.100	25.900	27.600	1010.100	80.300	6.800
6	28.600	28.400	4.600	24.200	20.800	1010	80.300	4.800
7	28.800	28.700	2.800	24.800	33.800	1008.700	80.800	6.200
8	28.100	28.021	1.800	24.800	2.800	1011	82.600	0
9	28.600	28.475	1.900	25	12.100	1010.200	81.200	7.300
10	28.800	28.743	1	24.400	4.800	1012.200	73.800	0.200
11	27.200	27.200	0.800	25.100	0	1012	80.200	7.300
12	28.700	28.367	1.700	24.800	0	1011.800	79.500	9.700
13	29.700	29.450	2	25.000	0	1010.600	78.100	6.400
14	30.300	30.133	1.700	25.800	0	1007.100	79.800	9.400
15	28.100	28	1.100	24.200	0	1007.600	79.400	0
16	27.600	27.300	1.200	24.800	1.100	1007.900	88	6.300

Gambar 4 Hasil Prediksi dengan Decision Tree

Tabel 2 Performace Decision Tree

No.	Uji Performance	Nilai
1.	Root Mean Squared Error (RMSE)	0.293 +/- 0.000
2.	Mean Absolute Error (MAE)	0.189 +/- 0.224
3.	Squared Correlation (R ²)	0.948

Performa model machine learning dari algoritma regresi linier pada tabel 1 menunjukkan hasil nilai RMSE sebesar 0,131 +/- 0,000, nilai MAE 0,099 +/- 0,086 dan nilai R² 0,990 sedangkan dari algoritma decision tree yang ditunjukkan tabel 2 menghasilkan nilai RMSE sebesar 0,293 +/-0,000, nilai MAE 0,189 +/-0,224 dan R² 0,948.

4. KESIMPULAN

Berdasarkan hasil analisis maka disimpulkan bahwa algoritma regresi linier dan algoritma *decision tree* pada *machine learning* dapat digunakan untuk memprediksi temperatur di Kota Semarang.

Dari hasil analisis yang didapatkan algoritma *regresi linier* menghasilkan nilai RMSE sebesar 0,131 +/- 0,000, nilai MAE 0,099 +/- 0,086 dan nilai R² 0,990, sedangkan algoritma *decision tree* menghasilkan nilai RMSE sebesar 0,293 +/-0,000, nilai MAE 0,189

+/-0,224 dan R² 0,948, sehingga algoritma *regresi linier* lebih akurat dibanding dengan *decision tree* karena dilihat dari korelasinya yang lebih mendekati 1 (satu) yaitu R² sebesar 0.990 dan tingkat error nya yaitu RMSE sebesar 0.131 +/- 0.000 dan MAE sebesar 0.099 +/- 0.086 lebih mendekati 0 (nol).

DAFTAR PUSTAKA

- [1] Citra Agustina Rahayu et al, "Prediksi Penderita Diabetes menggunakan metode Naive Bayes" JITET (Jurnal Informatika dan Teknik Elektro Terapan) Vol. 11 No. 3, pISSN: 2303-0577eISSN:2830-7062DOI: <https://doi.org/10.23960/jitet.v11i3.3055L>.
- [2] Aryani, Fatmasari, Afriyudi, and N.Hadinata, "Prediksi jumlah siswa baru dengan menggunakan metode exponential smoothing (studi kasus: SMK Ethika Palembang)," Bina Darma Conf.Comput....,pp.237–44,2020, [Online]. Available:<https://conference.binadarma.ac.id/index.php/BDCCS/article/view/1615%0Ahttps://conference.binadarma.ac.id/index.php/BDCCS/article/download/1615/693>
- [3] R. Limehuwey, W. Multi, and S. H. Kotarumalos, "Prediksi Suhu Udara Rata-rata Harian Dengan Metode Regresi Linier di Kota Ambon," *Tanah Goyang*, vol. 1, no. 1, pp. 18–25, 2023.
- [4] R. Akbar, R. Santoso, and B. Warsito, "Prediksi Tingkat Temperatur Kota Semarang Menggunakan Metode Long Short-Term Memory (Lstm)," *Jurnal Gaussian*, vol. 11, no. 4, pp. 572–579, 2023, doi: 10.14710/j.gauss.11.4.572-579.
- [5] Yuwono, B. D., Abidin, H. Z., & Hilmi, M. (2013). "Analisa geospasial penyebab penurunan muka tanah di Kota Semarang". *Prosiding Sains Nasional dan Teknologi*, 1(1).
- [6] Ahmad Suryadi. "Tropical Rainfall and Its Effect on Urban Life in Indonesia". Springer, 2021 ISBN: 978-981-15-9476-4.
- [7] Yufahri, M. O., & Widjajanti, R. (2022). "Kajian Lahan Terbangun pada Kawasan Rawan Bencana Longsor di Kecamatan Tembalang, Kota Semarang". *Teknik PWK (Perencanaan Wilayah Kota)*, 11(3), 224-237.
- [8] R. Risanti, "Analisis Model Prediksi Cuaca Menggunakan Support Vector Machine, Gradient Boosting, Random Forest, Dan Decision Tree," vol. XII, pp. 119–128, 2024, doi: 10.21009/03.1201.fa18.
- [9] C. D. Usman, A. P. Widodo, K. Adi, and R. Gernowo, "Rainfall prediction model in

- Semarang City using machine learning,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 2, p. 1224, May 2023, doi: 10.11591/ijeecs.v30.i2.pp1224-1231.
- [9] Agusti Frananda Alfonsus Naibaho dan Amalia Zahra “Prediksi kelulusan siswa Sekolah Menengah Pertama menggunakan *Machine Learning*” *JITET (Jurnal Informatika dan Teknik Elektro Terapan)* - Vol. 11 No. 3, pISSN: 2303-0577 eISSN: 2830-7062 DOI: <https://doi.org/10.23960/jitet.v11i3.3056> -B. Wu, C. Zheng, “An Analysis of the Effectiveness of Machine Learning Theory in the Evolution of Education and Teaching”, *Hindawi Journal, Northeast Normal University (China)*, October 11, 2021, pp. 1-10.
- [10] F. Indriaharti Harida and N. Khazizah, “Analisis Cuaca Di Kota Jakarta Bulan Januari Tahun 2018 Menggunakan Algoritma Decision Tree,” *Jurnal Poros Teknik*, vol. 14, no. 1, pp. 33–37, 2022, [Online]. Available: <https://www.kaggle.com/datasets/msf1203/pr>
- [11] K. Halim, D. Erny Herwindiati, and T. Sutrisno, “Penerapan Metode Decision Tree Untuk Prakiraan Cuaca Kota Bekasi,” *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 11, no. 2, pp. 1–5, 2023, doi: 10.24912/jiksi.v11i2.26026.
- [12] S. Hidayatullah and A. Cherid, “Prediksi Temperatur Cuaca di Negara Norwegia Menggunakan Metode LSTM,” *Simkom*, vol. 8, no. 2, pp. 187–198, 2023, doi: 10.51717/simkom.v8i2.192.
- [13] I. Intan, S. Aminah Dinayati Ghani, A. T. Koswara, U. Dipa Makassar, K. Arsip Nasional Republik Indonesia, and J. P. Kemerdekaan, “Analisis Performansi Prakiraan Cuaca Menggunakan Algoritma Machine Learning Performance Analysis of Weather Forecasting using Machine Learning Algorithms,” *Jurnal Pekommas_Vol._6_No*, vol. 2, pp. 1–8, 2021, doi: 10.30818/jpkm.2021.2060221.
- [14] A. Y. Barrera-Animas, L. O. Oyedele, M. Bilal, T. D. Akinosho, J. M. D. Delgado, and L. A. Akanbi, “Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting,” *Machine Learning with Applications*, vol. 7, no. August 2021, p. 100204, 2022, doi: 10.1016/j.mlwa.2021.100204.
- [15] D. A. Widodo, N. Iksan, E. D. Udayanti, and Djuniadi, “Renewable energy power generation forecasting using deep learning method,” *IOP Conf Ser Earth Environ Sci*, vol. 700, no. 1, 2021, doi: 10.1088/1755-1315/700/1/012026.
- [16] Djuniadi *et al.*, “Face mask detection services of Covid19 monitoring system to maintain a safe environment using deep learning method,” *IOP*