



Analisis Ketercapaian Vaksinasi terhadap Penyebaran COVID-19 Menggunakan *Machine Learning*

Syafrial Fachri Pane^{1*}, Ferdy Berliano Putra², Gilang Romadhanu Tartila³, Chandra Ahmad Rizki⁴

^{1,2}*Teknik Informatika, Politeknik Pos Indonesia, Jl. Sari Asih No.54, Bandung 4151, Indonesia*

^{3,4}*Teknik Informatika, Universitas Esa Unggul Jl. Arjuna Utara No.9, Jakarta Barat, Indonesia*

*Email Penulis Koresponden: syafrial.fachri@poltekpos.ac.id

Abstrak:

Virus COVID-19 merupakan virus yang muncul pada akhir tahun 2019 yang menyerang saluran pernafasan dan menyebar dengan cepat sehingga menyebabkan dampak pandemik secara global. Untuk mencegah COVID-19 dilakukan beberapa cara diantaranya yaitu menjaga jarak, mencuci tangan, dan menggunakan masker namun tidak mengurangi penyebaran COVID-19. Untuk mengurangi dampak lebih lanjut dari penyebaran COVID-19 dibuat vaksin untuk memperkuat sistem kekebalan tubuh. Untuk melihat ketercapaian vaksinasi di suatu daerah diperlukan alat yang dapat bekerja secara otomatis yang bekerja dari data tanpa pemrograman eksplisit menggunakan *Machine Learning*. Adapun data cakupan vaksinasi yang diprediksi adalah pada provinsi Jakarta dengan sumber melalui website Satgas COVID-19 dengan rentang waktu dari tanggal 1 Juli 2021 hingga 30 November tahun 2021 dengan parameter yang akan diuji adalah sasaran, belum vaksin, dosis 1, dosis 2, total vaksin diberikan. Model yang diusulkan dalam penelitian kali ini yaitu *AdaBoost Regressor*. Kinerja *regressor* ditentukan berdasarkan akar rata-rata keasalahan (RMSE) dan kesalahan mutlak (MAE). Setelah melakukan pengujian menggunakan model maka di dapatkan nilai keakuratan terhadap model maka nilai yang didapatkan adalah 98%. Berdasarkan berita dari Kementerian Kesehatan Indonesia dikatakan tercapainya vaksinasi jika sudah mencapai lebih kurang 80% untuk total vaksinasi yang sudah diberikan. Total dosis vaksinasi yang sudah disebarkan pada masyarakat di provinsi DKI Jakarta dengan Kabupaten Adm. Kep. Seribu sebagai kota/kabupaten yang tertinggi yang sudah mencapai 140% total dosis yang diberikan.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license



Kata Kunci:

Vaksinasi; AdaBoostRegressor; Machine Learning; COVID-19;

Riwayat Artikel:

Diserahkan 20 April 2022
Direvisi 25 Agustus 2022
Diterima 9 September 2022
Dipublikasi 26 Desember 2022

DOI:

10.22441/incomtech.v12i3.15370

1. PENDAHULUAN

Virus merupakan penyakit menular yang sudah teridentifikasi dari pertengahan tahun 1900-an yang tidak hanya mengancam kesehatan namun juga kesehatan secara global yang mana 60% penyakit manusia yang muncul teridentifikasi dari bersifat zoonosis, yang artinya virus endemik dalam populasi hewan yang menyebar melalui berbagai vektor dan jalur transmisi lainnya hingga terpapar kepada manusia [1]. Pada abad ke -21 ini virus penyakit menular juga menjadi tantangan, sebagai contohnya adalah Ebola dan sindrom pernafasan Timur Tengah yang menjadi masalah besar dalam kesehatan [2].

Pada akhir tahun 2019 tepatnya bulan Desember Organisasi Kesehatan Dunia (WHO) mengatakan virus baru bernama COVID-19 ditemukan di daerah Wuhan, Cina dengan cepat mempengaruhi masyarakat setempat dan juga negara Cina, dengan cara kerja virus ini yaitu tertular melalui pernafasan dan dalam kurun waktu beberapa bulan virus ini menyerang semua negara dan ditetapkan sebagai pandemi global [2]. Virus COVID-19 ini juga alias *Severe Acute Respiratory Syndrome CoronaVirus-2*. COVID-19 termasuk ke dalam sekelompok penyakit yang mengakibatkan seperti batuk dan influenza yang dipicu oleh *Middle East Respiratory Syndrome CoronaVirus* [3]. Virus corona menyebar dengan sangat cepat, pada bulan Februari 2020 tercatat sudah 3,5 juta orang terpapar kasus ini di dunia dengan Amerika sebagai negara terinfeksi mencapai 1,1 juta [4]. Ini membuktikan dengan hanya memakai masker, pembatasan jarak, membersihkan tangan menggunakan sabun tidak menurunkan resiko terdampak virus COVID-19. Oleh karena itu WHO menyetujui vaksinasi dan tersedia untuk seluruh masyarakat di dunia guna meningkatkan kekebalan tubuh manusia supaya tidak mudah terinfeksi oleh virus COVID-19 [5].

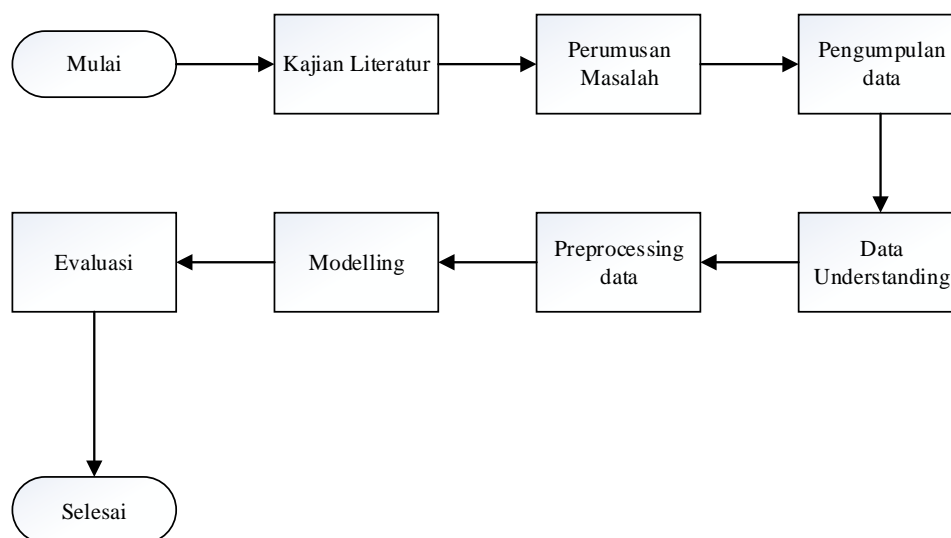
Dari permasalahan diatas maka dapat diambil kesimpulan pada penelitian ini akan melakukan analisis ketercapaian vaksinasi di suatu wilayah tertentu di Indonesia. Vaksinasi diharapkan dapat memperkuat sistem kekebalan tubuh masyarakat dan kembali ke kehidupan normal. Untuk melihat perkembangan vaksinasi diperlukan alat yang dapat bekerja secara otomatis yang bekerja dari data tanpa pemrograman eksplisit yaitu menggunakan *machine learning*. *Machine learning* adalah alat yang digunakan untuk penelitian yang membantu komputer dalam membuat pemodelan melalui pengalaman dan dapat memprediksi peristiwa di masa depan [6]. Pada *machine learning* terdapat kategori *supervised learning* yaitu pembelajaran secara statistik yang menghubungkan keluaran model yang diinginkan dengan masukannya, hubungan pembelajaran dari data berlabel, yaitu data di mana baik masukan (prediktor, variabel bebas) dan keluaran (tanggapan, variabel terikat) diketahui [7]. Hal tersebut dapat dibagi menjadi dua metode yaitu regresi dan klasifikasi. Teknik regresi digunakan untuk menggambarkan item data dengan prediksi aktual. Regresi merupakan suatu model pada *supervised learning* atau pembelajaran yang diawasi untuk permasalahan dalam analisis regresi di mana variabel yang menarik mengasumsikan nilai dalam interval biner [8]. Tujuan utama dari metode regresi adalah untuk menarik garis atau kurva yang paling cocok dari data.

Pada penelitian ini, data cakupan vaksinasi yang diprediksi adalah pada provinsi DKI Jakarta dengan parameter yang akan di uji adalah sasaran, belum vaksin, dosis 1, dosis 2, total vaksin diberikan. Berdasarkan jenis data *supervise learning* maka model yang digunakan adalah *regressor* yaitu *AdaBoostRegressor* [9]. Kinerja *regressor* ditentukan berdasarkan *Root Mean Squared Error* atau akar rata-rata keasalahan (RMSE) dan *Mean Absolute Error* atau kesalahan mutlak (MAE) [10]. MAE dan RMSE merupakan metrik yang memperkirakan kuantitas yang sama dalam faktor skala [11]. Semakin rendah nilai RMSE, maka semakin tinggi efisiensi model digunakan. Berikut penghitungan RMSE dan MAE.

2. METODE PENELITIAN

Pada tahapan ini hal yang penulis gunakan dalam metodologi penelitian adalah menggunakan panduan dari metode CRISP-DM. Metode ini yang sering dipergunakan dan juga sering digunakan untuk suatu perusahaan.

Dari metodologi penelitian diatas maka langkah-langkah penelitian yang dilakukan dapat digambarkan pada Gambar 1.



Gambar 1. Alur Metodologi Penelitian

Gambar 1 merupakan langkah-langkah metodologi yang dimulai dari mencari referensi berkaitan dengan penelitian untuk memperoleh hasil diharapkan dari penelitian yang dilakukan. Kemudian ditentukan identifikasi masalah atau perumusan masalah sebagai rancangan permasalahan pada penelitian. Setelah didapatkan permasalahan yang didapatkan kemudian dikumpulkan data yang bersumber dari website satgas COVID-19 DKI Jakarta dengan rentang waktu pengambilan data 1 Juli 2021 hingga 30 November 2021. Setelah didapatkan data yang akan digunakan melakukan analisa terhadap data untuk dapat membangun data dan membuat model terbaik untuk data.

Selanjutnya melakukan *preprocessing* data yang untuk dilanjutkan ke dalam penelitian dengan tahapan diantaranya yaitu data *cleansing*, *missing values* dan transformasi. Pada data *pre-procesing* ini peneliti melakukan tahap *cleaning data*

yaitu dengan cara membersihkan data dengan cara *outlier*. *Outlier* adalah data yang menyimpang dari objek lainnya yang dapat menimbulkan keraguan terhadap objek data. Adapun metode *outlier* yang digunakan adalah dengan metode IQR (*Interquartile range*) [12]. Kemudian untuk tahap *missing values* atau nilai yang kosong peneliti menggunakan metode interpolasi linier [13]. Interpolasi linier adalah suatu metode interpolasi untuk suatu dimensi pada data. Dengan memperkirakan nilai data berdasarkan dua titik data yang berdekatan dengan titik yang perlu diinterpolasi dalam urutan data satu dimensi [14].

Setelah itu dilakukan tahap metode statistika dan *machine learning* untuk penentuan *modelling* data yang diterapkan pada *dataset* yang telah tersedia. Dalam penelitian ini menggunakan metode yaitu *AdaBoostRegressor*. Selanjutnya model yang digunakan kinerja model *AdaBoostRegressor* dievaluasi menggunakan metode berdasarkan *Root Mean Squared Error* atau akar rata-rata keasalahan (RMSE) dan *Mean Absolute Error* atau kesalahan mutlak (MAE) [15].

3. HASIL DAN PEMBAHASAN

Untuk penelitian ini *dataset* yang digunakan dikumpulkan melalui website satgas COVID-19 DKI Jakarta dari tanggal 1 Juli sampai 30 November tahun 2021 (Tabel 1). Data yang digunakan yaitu data perkelurahan di provinsi DKI Jakarta yaitu 267 kelurahan 44 kecamatan 5 Kotamadya dan 1 Kabupaten. Data yang digunakan bersifat kumulatif yang dikumpulkan perhari. Atribut yang digunakan yaitu “SASARAN” yaitu sasaran atau target berapa banyak penduduk yang akan di vaksin nantinya. Kemudian attribute selanjutnya yaitu “BELUM VAKSIN” yaitu berapa sisa warga yang belum di vaksin berdasarkan target atau sasaran. Lalu atribut digunakan yaitu “DOSIS 1” yaitu jumlah vaksin dosis 1 yang sudah diberikan kepada penduduk, “DOSIS 2” yaitu jumlah vaksin dosis 2 yang sudah diberikan kepada penduduk, “TOTAL” yaitu penjumlahan dosis vaksin 1 dan vaksin dosis 2 yang sudah di berikan pada penduduk.

Tabel 1. *Dataset*

Tanggal	Kota	Kecamatan	Kelurahan	Sasaran	Belum Vaksin	Dosis 1	Dosis 2	Total
2021/07/01	Jakarta Utara	Pademangan	Ancol	20393	13272	7114	3287	10401
2021/07/01	Jakarta Barat	Tambora	Angke	25785	16477	9299	3221	12520
2021/07/01	Jakarta Timur	Kramat Jati	Bale Kambang	25158	18849	6301	2644	8945
2021/07/01	Jakarta Timur	Jatinegara	Bali Mester	8683	5743	2937	1517	4454
2021/07/01	Jakarta Timur	Cipayung	Bambu Apus	22768	15407	7357	3985	11342

Data yang digunakan dalam penelitian ini pada waktu tertentu memiliki data yang kosong, sehingga diisi menggunakan metode interpolasi linier. Adapun rumus yang dapat digunakan untuk perhitungan menggunakan interpolasi linier.

$$y(x) = y_1 + (x - x_1) \frac{(y_2 - y_1)}{(x_2 - x_1)} \quad (1)$$

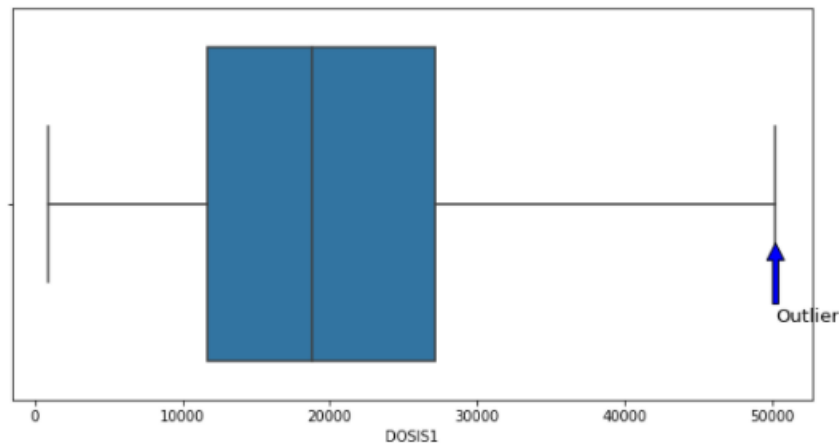
x_1, y_1 adalah koordinat titik pertama

x_2, y_2 adalah koordinat titik kedua

x adalah titik dimana dilakukan interpolasi

y adalah nilai interpolasi

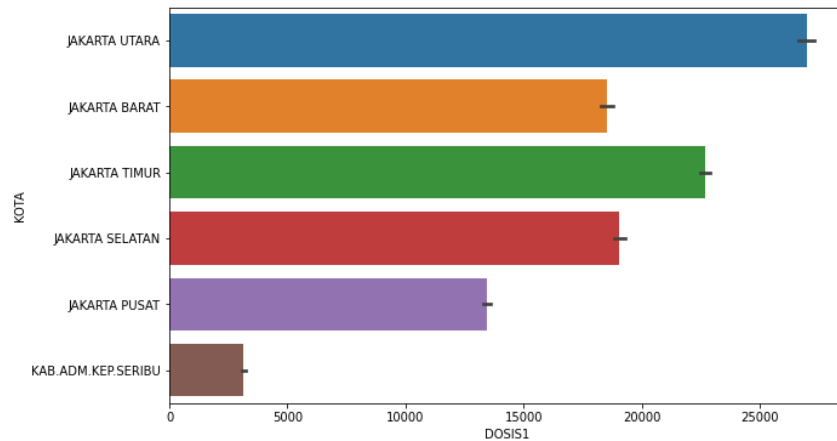
Data kemudian dibersihkan dengan menggunakan cara *outlier* seperti yang ditunjukkan pada [Gambar 2](#). *Outlier* sendiri yaitu objek yang menyimpang dari objek lainnya. Metode *outlier* yang digunakan pada penelitian ini adalah metode *inter-quartile range* (IQR). Untuk menentukan metode IQR sendiri yaitu dengan cara menentukan batas bawah dan batas atas. Untuk menentukan batas bawah yaitu mencari tahu percentil dari 25% atribut yang digunakan, sedangkan untuk batas atas dengan mencari tahu dari percentil 75% atribut yang digunakan. Jika sudah didapatkan maka hitung IQR dengan cara mengurangi nilai batas atas dengan batas bawah. Setelah didapatkan kita melanjutkan untuk membuat nilai batas bawah dengan batas bawah dikurangi 1,5 dikalikan dengan nilai IQR, dan batas atas dengan batas atas ditambah 1,5 dikali IQR maka didapatkanlah nilai batas atas dan batas bawahnya.



Gambar 2. Pembersihan *Outliers*

Setelah selesai membersihkan data, lihat ketercapaian vaksinasi berdasarkan wilayah kota. Capaian vaksinasi perwilayah kota dapat dilihat pada [Gambar 3](#).

Setelah selesai menyiapkan data, peneliti menentukan pemodelan yaitu *regressor* dengan teknik *AdaBoostRegressor*. Kemudian dilakukan pemisahan *dataset* menjadi 2 yaitu *train* dan *test set* dengan perbandingan 70:30. Pembagian untuk *train set* 70% dan untuk pembagian *test set* 30%. Setelah selesai menerapkan model maka cari nilai yang didapatkan setelah menerapkan model tersebut. Nilai yang didapatkan adalah 98% dengan nilai korelasi 99%.



Gambar 3. Capaian Vaksinasi Perwilayah Kota

Setelah selesai melakukan pemodelan maka dilakukan peninjauan ulang terhadap model dengan menggunakan MAE dan RMSE. Adapun rumus yang dapat digunakan untuk perhitungan menggunakan MAE dan RMSE sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad (2)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (3)$$

e_t = nilai aktual

n = banyaknya data

\sum = Jumlah keseluruhan nilai

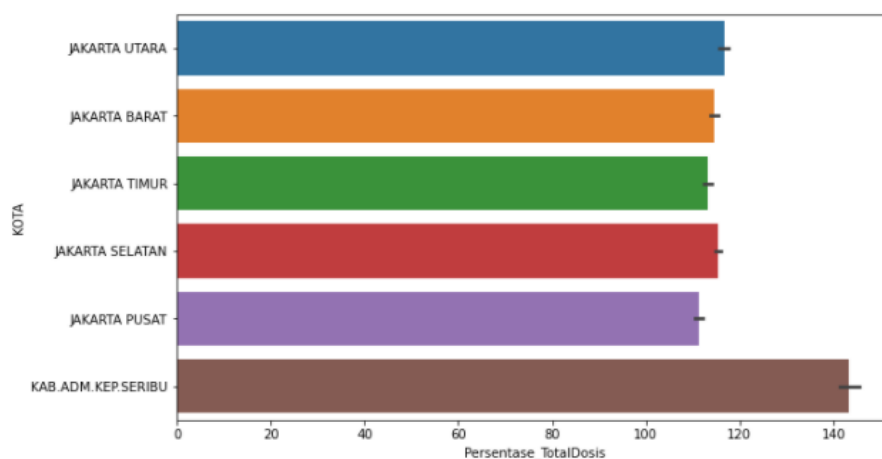
Dengan demikian didapatkan nilai evaluasi MAE dan RMSE seperti yang ditunjukkan pada [Tabel 2](#) di bawah ini.

Tabel 2. Nilai Evaluasi MAE dan RMSE

Model	MAE	RMSE
<i>AdaboostRegressor</i>	1040.36	1240.83

Berdasarkan info Kementerian Kesehatan Indonesia dikatakan tercapai vaksinasi adalah jika vaksinasi sudah mencapai lebih kurang 80% dosis yang sudah disebarkan baik dosis1 dan dosis 2.

Pada [Gambar 4](#) bisa dilihat total dosis vaksinasi yang sudah disebarkan pada masyarakat di provinsi DKI Jakarta dengan Kabupaten Adm. Kep. Seribu sebagai kota/kabupaten yang tertinggi untuk memberikan vaksinasi dosis 1 dan dosis 2 yang sudah mencapai 140% total dosis yang diberikan.



Gambar 4. Persentase Total Dosis Vaksinasi

4. KESIMPULAN

Untuk mencegah penyakit yang sangat menular dari COVID-19, proses vaksinasi yang aman dan efektif adalah wajib. Proses vaksinasi dapat menghilangkan rasa takut terinfeksi saat terpapar dengan orang lain. Pada penelitian Analisis Ketercapaian Vaksinasi Terhadap Penyebaran COVID-19 di DKI Jakarta menggunakan model *AdaBoost Regressor*. Sebelum menerapkan model data dibersihkan dengan metode *outlier* yang di cek menggunakan *boxplot* dengan metode IQR (*Inter-quartile range*) lalu data di uji menggunakan model *Adaboost* dengan nilai akurasi senilai 98% dan nilai korelasi 99%. Namun sangat disayangkan nilai dari MAE dan RMSE masih jauh dari nilai nol yaitu MAE dengan nilai yang didapatkan yaitu 1040.3676 dan nilai RMSE yaitu 1240.8315. Melalui info Kementerian Kesehatan Indonesia tercapainya vaksinasi jika sudah mencapai lebih kurang 80% untuk total vaksinasi yang sudah diberikan yaitu baik vaksinasi dosis 1 dan dosis 2.

REFERENSI

- [1] S. Balinandi *et al.*, "Identification and molecular characterization of highly divergent RNA viruses in cattle, Uganda," *Virus Res.*, vol. 313, no. March, p. 198739, 2022, doi: 10.1016/j.virusres.2022.198739.
- [2] S. Yazdanirad, M. Sadeghian, M. Jahadi Naeini, M. Abbasi, and S. M. Mousavi, "The contribution of hypochondria resulting from Corona virus on the occupational productivity loss through increased job stress and decreased resilience in the central workshop of an oil refinery: A path analysis," *Heliyon*, vol. 7, no. 4, p. e06808, 2021, doi: 10.1016/j.heliyon.2021.e06808.
- [3] P. Wu *et al.*, "Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020," *Eurosurveillance*, vol. 25, no. 3, pp. 1–6, 2020, doi: 10.2807/1560-7917.ES.2020.25.3.2000044.
- [4] M. C. Younis, "Evaluation of deep learning approaches for identification of different coronavirus species and time series prediction," *Comput. Med. Imaging Graph.*, vol. 90, no. January, p. 101921, 2021, doi: 10.1016/j.compmedimag.2021.101921.
- [5] A. Khakimova, L. Abdollahi, O. Zolotarev, and R. Fakher, "Global interest in vaccines during the COVID-19 pandemic: evidence from Google Trends," *Vaccine X*, vol. 10, p. 100152, 2022, doi: 10.1016/j.jvax.2022.100152.
- [6] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," *Expert Syst. Appl.*, vol. 166, p. 114060, 2021, doi: 10.1016/j.eswa.2020.114060.

- [7] J. Schmidt, M.R.G. Marques, S. Botti. et al. Recent advances and applications of machine learning in solid-state materials science. *npj Comput Mater*, vol. 5, p.83, 2019, doi: 10.1038/s41524-019-0221-0.
- [8] G. Nicora, M. Rios, A. Abu-Hanna, and R. Bellazzi, "Evaluating pointwise reliability of machine learning prediction," *J. Biomed. Inform.*, vol. 127, p. 103996, 2022, doi: 10.1016/j.jbi.2022.103996.
- [9] X. Zhu, P. Zhang, and M. Xie, "A Joint Long Short-Term Memory and AdaBoost regression approach with application to remaining useful life estimation," *Meas. J. Int. Meas. Confed.*, vol. 170, no. September, p. 108707, 2021, doi: 10.1016/j.measurement.2020.108707.
- [10] S. Dutta, U. Mukherjee, and S. Kumar Bandyopadhyay, "Pharmacy Impact on Vaccination Progress Using Machine Learning Approach," *Preprints*, 2021, doi: 10.20944/preprints202106.0533.v1.
- [11] D. S. K. Karunasingha, "Root mean square error or mean absolute error? Use their ratio as well," *Inf. Sci. (Ny)*, vol. 585, pp. 609–629, Mar. 2022, doi: 10.1016/J.INS.2021.11.036.
- [12] N. N. R. R. Suri, N. Murty, and M. G. Athithan, *Outlier Detection: Techniques and Applications - A Data Mining Perspective*. 2019.
- [13] B. Ouyang, Y. Song, Y. Li, G. Sant, and M. Bauchy, "EBOD: An ensemble-based outlier detection algorithm for noisy datasets," *Knowledge-Based Syst.*, vol. 231, p. 107400, 2021, doi: 10.1016/j.knosys.2021.107400.
- [14] G. Huang, "Missing data filling method based on linear interpolation and lightgbm," *J. Phys. Conf. Ser.*, vol. 1754, no. 1, 2021, doi: 10.1088/1742-6596/1754/1/012187.
- [15] R. M.-G. Jesús de-Prado-Gil, Covadonga Palencia, Neemias Silva-Monteiro, "To predict the compressive strength of self compacting concrete with recycled aggregates utilizing ensemble machine learning models," *Case Stud. Constr. Mater.*, vol. 16, no. April, p. e01046, 2022, doi: 10.1016/j.cscm.2022.e01046.