

PERBANDINGAN HASIL PREDIKSI HARGA PROPERTI DI DAERAH BROOKLYN MENGGUNAKAN METODE XGBOOST, RANDOM FOREST, DAN LINEAR REGRESSION

Adyatma Ruliff Brahmantyo

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
E-mail: adyatma.535210047@stu.untar.ac.id,

ABSTRAK

Hukum supply and demand menyebabkan kenaikan harga properti karena permintaan yang terus meningkat untuk tempat tinggal di daerah padat penduduk seperti Brooklyn, New York. Tujuan dari penelitian ini adalah untuk menemukan model terbaik berdasarkan efisiensi dan akurasi dengan menggunakan algoritma pengajaran mesin seperti *Linear Regression*, Random Forest, dan XGBoost untuk memprediksi harga properti. Dataset yang digunakan terdiri dari 20.894 data penjualan properti yang memiliki atribut seperti harga penjualan, luas bangunan, dan jumlah unit. Dengan menggunakan random_state acak, eksperimen dilakukan dengan membagi data latihan dan uji menjadi 70:30. Memanfaatkan metrik R², MAE, RMSE, dan waktu komputasi, kinerja algoritma dinilai. Hasil menunjukkan bahwa XGBoost memiliki kinerja terbaik dengan nilai R² tertinggi (0.014) dan MAE dan RMSE terendah. Di sisi lain, Linear Regression menunjukkan bahwa XGBoost memiliki kinerja terburuk dengan nilai R² negatif dan eror yang tinggi. Penelitian ini menemukan bahwa XGBoost memiliki kemampuan yang lebih baik untuk memodelkan data properti dan memberikan prediksi yang lebih akurat daripada metode lain. Hasil-hasil ini dapat membantu pemangku kepentingan, seperti pengembang properti dan pembeli, memahami tren harga properti dan membuat keputusan yang lebih baik berdasarkan data.

Kata kunci—Prediksi, *Machine Learning*, XGBoost, *Linear Regression*, *Random Forest*,

ABSTRACT

Due to the law of supply and demand, real estate prices are rising in highly populated places like Brooklyn, New York, as a result of the growing demand for homes. In order to determine which machine learning algorithms—Random Forest, XGBoost, and Linear Regression—perform best in terms of accuracy and efficiency, this study attempts to forecast real estate values. 20,894 property sales records, comprising details such as building area, number of units, sale price, and more, make up the dataset that was used. Using randomized seeds, the data was divided into training and testing sets (70:30) for the experiment. R², MAE, RMSE, and computation time measures were used to assess the algorithms' performance. With the lowest MAE and RMSE and the greatest R² value (0.014), the findings demonstrated that XGBoost performed better than the other methods. On the other hand, with negative R² values and large error rates, Linear Regression performed the worst. This study shows that XGBoost outperforms other approaches in modeling property data, yielding more accurate predictions. The results can help stakeholders, including purchasers and developers, comprehend patterns in real estate prices and make data-driven

Keywords—Prediction, *Machine Learning*, XGBoost, *Linear Regression*, *Random Forest*

1. PENDAHULUAN

Dewasa ini permintaan akan tempat tinggal sangat tinggi, terutama pada daerah yang memiliki tingkat kepadatan penduduk yang tinggi seperti negara bagian Amerika Serikat, yaitu New York. Berdasarkan data yang dipublikasikan oleh sensus amerika serikat total jumlah penduduk negara bagian New York pada tahun 2020 sebanyak 20 juta jiwa [1]. Terkhusus pada daerah Brooklyn terdapat 2.7 juta penduduk [2]. Data populasi tersebut berdasarkan hasil sensus pada tanggal 1 Juli 2020. Semakin meningkatnya jumlah penduduk maka permintaan terhadap tempat tinggal akan semakin tinggi, hal ini akan membuat harga properti semakin tinggi [3], [4]. New York merupakan tempat yang mirip dengan Jakarta, di mana banyak orang yang berbondong-bondong pindah untuk

mengadu nasib. Hal ini berlaku hukum *supply and demand* di mana terjadinya kenaikan harga rumah, di mana suplai rumah yang terbatas dan permintaan terhadap rumah semakin tinggi [5].

Penelitian ini akan melakukan prediksi menggunakan machine learning. Machine learning adalah sebuah cabang dari ilmu komputer yang bertujuan untuk mengajari komputer memprediksi, melakukan klasifikasi, dan lain-lain [6]. Machine learning dapat meningkatkan pemahaman tentang analisis dan mengurangi kompleksitas penelitian. Machine learning juga dapat meningkatkan akurasi prediksi [6], [7], [8]. Setelah dilakukannya eksperimen perlu dievaluasi untuk mengukur tingkat akurasi dari model machine learning [9]. Dalam penelitian ini akan membandingkan tiga metode machine learning, yaitu *XGBoost* [10], *Random Forest* [11], dan *Linear Regression* [12]. Metode tersebut sangat umum digunakan untuk melakukan prediksi. Metrik pembanding tiga algoritma tersebut adalah kecepatan komputasi, R^2 [13], *mean absolute error* (MAE) [14], dan *root means square error* (RMSE) [15]. R^2 adalah sebuah ukuran statistik yang menunjukkan seberapa baik tidaknya sebuah model regresi [16]. MAE adalah sebuah matrik yang mengukur rata-rata absolut antara nilai prediksi dan nilai aktual atau nilai asli [17]. Sedangkan RMSE adalah sebuah metrik yang mengukur akar kuadrat dari rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual [17]. Dan yang terakhir adalah waktu komputasi, Di mana kode akan dihitung saat menjalankan dan melatih *machine learning* sampai selesai melatih dan memprediksi nilai.

Data yang digunakan didapatkan dari pemerintah negara bagian New York. Data yang digunakan adalah data penjualan properti yang ada di daerah Brooklyn. Banyak data yang ada adalah 20.894 data. *Borough, neighborhood, building class category, tax class at present, block, lot, easement, building class at present, address, apartment number, zip code, residential units, commercial units, total units, land square feet, gross square fee, year built, tax class at time of sale, building class at time of sale, sale price, sale date* merupakan kolom yang terdapat pada data yang digunakan.

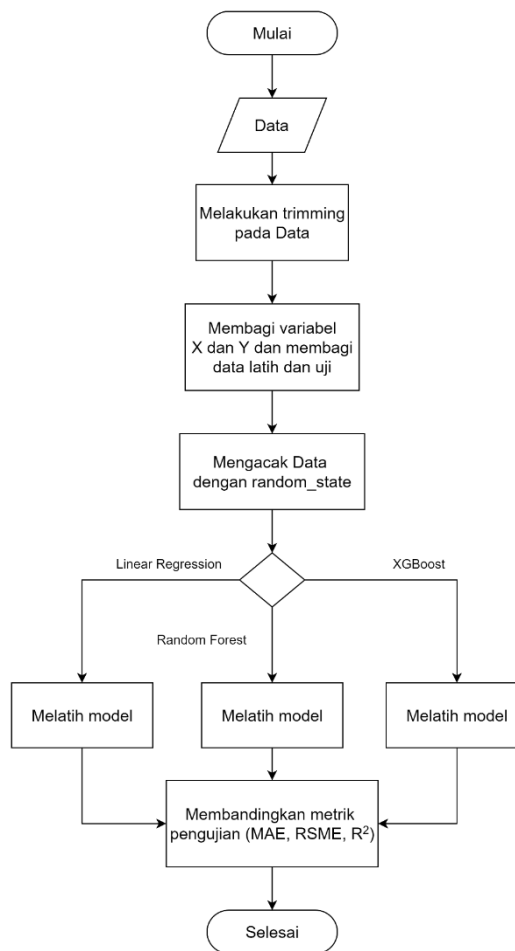
2. METODE PENELITIAN

Skenario pelaksanaan penelitian dapat disebut juga sebagai metode penelitian [18]. Ada banyak metode penelitian, di antaranya adalah kuantitatif, kualitatif, eksperimen, dan lain-lain. Dalam penelitian ini metode yang digunakan adalah eksperimen. Metode eksperimen bertujuan untuk menjelaskan hubungan sebab-akibat (kausalitas) antara variabel tertentu [19].

2.1 Metode Penelitian

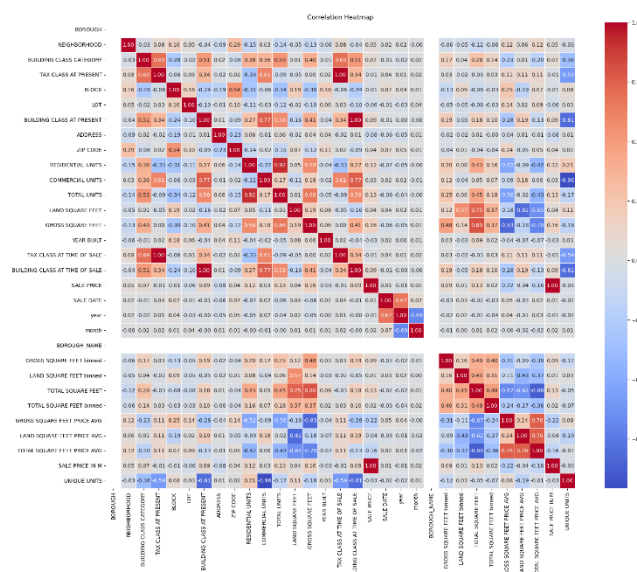
Tahapan dalam melakukan eksperimen pada penelitian ini dapat dilihat pada gambar 1. Pertama-tama adalah masukkan dataset ke dalam algoritma. kedua membagi variabel x dan y dan membagi data menjadi data latih dan data uji, dalam penelitian ini akan membandingkan data latih dan data uji, yaitu di angka 70:30. Kemudian mengacak data dengan menggunakan *random_state*. Ketiga data akan digunakan untuk melatih tiga model atau algoritma, yaitu regresi linear, *random forest*, dan *XGBoost*. Selama pelatihan model akan dihitung waktu komputasinya untuk mengetahui algoritma mana yang memiliki error paling sedikit, dan mendapatkan nilai R^2 yang paling tinggi.

Gambar 1 merupakan alur eksperimen pada penelitian ini akan dilakukan. Perhitungan untuk MAE, RMSE, dan R^2 diletakkan di akhir dan setelah perhitungan waktu komputasi ditunjukkan agar perhitungan tidak mengganggu waktu komputasi. Jadi akan perhitungan matrik pembanding akan dipisahkan dari pelatihan model. Parameter dalam eksperimen ini akan diacak dengan rentang 1 – 1000 untuk *random_state*.

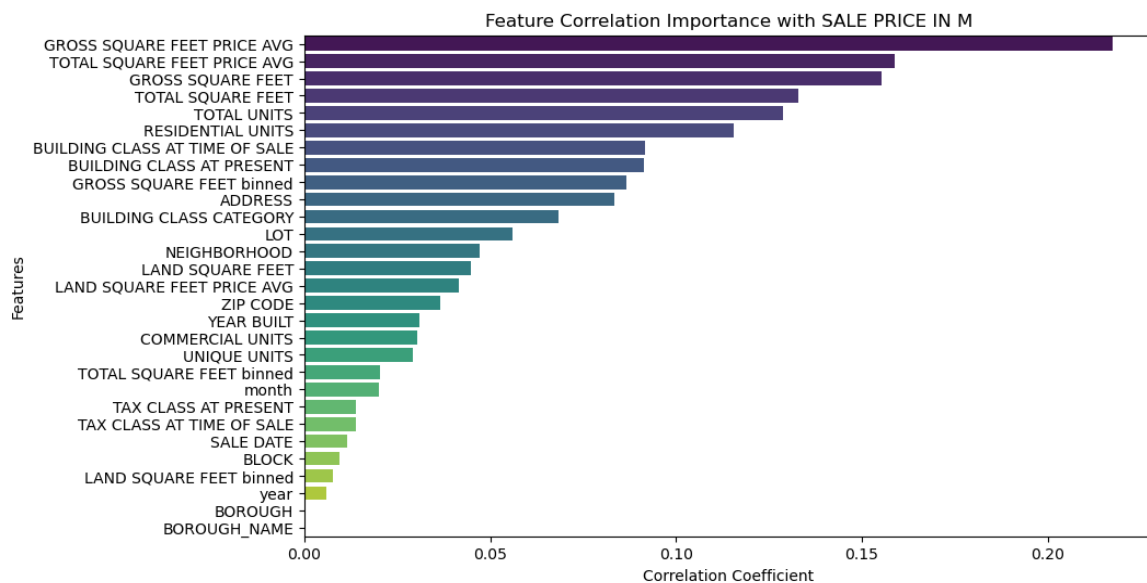


2.1.1 Data Input

Data yang digunakan dalam penelitian ini merupakan data time-series. Data time-series merupakan data yang berurutan berdasarkan waktu [20]. Sebelum dataset digunakan untuk pelatihan model dianalisis terlebih dahulu. Pertama-tama dapat dilihat korelasi antar data, korelasi antar data dapat dilihat pada gambar 1. Lalu setelah itu dilihat apa hal yang paling mempengaruhi harga jual, visualisasi dapat dilihat pada gambar 2.



Gambar 1 Korelasi antar data



Gambar 2 Visualisasi korelasi harga jual dengan kolom lain

Berdasarkan gambar 2 didapati bahwa *gross square feet price average* sangat berpengaruh terhadap harga jual.

2.1.2 Algoritma

Algoritma adalah kumpulan instruksi untuk memecahkan masalah [21]. Dalam penelitian ini akan membandingkan algoritma *linear regression*, *random forest*, dan *XGBoost*. Algoritma tersebut merupakan algoritma yang biasa digunakan untuk melakukan prediksi.

Linear regresi adalah suatu Teknik yang digunakan untuk menghubungkan linear antara variabel respons dengan satu atau lebih variabel independen [22]. Tujuan dari regresi linear adalah untuk memprediksi nilai variabel respons berdasarkan nilai variabel independen yang diketahui dengan memahami seberapa besar pengaruh variabel independen terhadap variabel respons.

Random forest adalah sebuah metode dalam *machine learning* yang berguna untuk membuat pohon keputusan untuk membuat model yang lebih akurat [23]. Jadi dalam *random forest* akan membuat banyak pohon keputusan secara acak. Prediksi akhir yang didapatkan adalah hasil mayoritas suara.

XGBoost (eXtreme Gradient Boosting) adalah pustaka pembelajaran mesin terdistribusi sumber terbuka yang menggunakan pohon keputusan, algoritma peningkatan pembelajaran terawasi yang menggunakan penurunan gradient [10]. Algoritma ini terkenal karena kecepatan, efisiensi, dan kemampuan untuk menskalakan data yang besar. *XGBoost* terus menerus melakukan iterasi selama proses pembelajaran. Pohon baru dihasilkan selama iterasi agar sesuai dengan residu pohon sebelumnya. Seiring dengan berjalannya iterasi, akurasi meningkat [24].

2.1.3 Metode Evaluasi

Dalam menentukan apakah sebuah algoritma atau model bagus atau tidak dapat ditentukan melalui metode evaluasi. Metode evaluasi yang digunakan dalam penelitian ini adalah R^2 , *mean absolute error* (MAE), dan *root means square error* (RMSE).

R^2 bertujuan untuk melakukan pengukuran proporsi varians dalam variabel dependen yang dijelaskan oleh variabel independen. Semakin tinggi nilai R^2 semakin tinggi maka semakin baik tingkat kecocokannya. Kegunaan dari R^2 adalah untuk membandingkan model yang berbeda pada data set yang sama. Namun dapat menyesatkan jika data tidak cocok untuk regresi [25].

MAE atau *mean absolute error* bertujuan untuk mengukur perbedaan absolut rata-rata antara nilai prediksi dan aktual. Semakin kecil nilai dari MAE maka semakin baik. Mae dapat digunakan untuk data dengan nilai yang ekstrem, mudah diinterpretasikan karena dalam unit yang sama dengan variabel dependen [25].

RMSE atau *root mean square error* bertujuan untuk memberikan kesalahan dalam unit yang sama dalam variabel dependen. RMSE semakin rendah maka menunjukkan akurasi yang lebih baik. Umumnya RMSE digunakan dalam analisis regresi karena sifatnya yang seimbang [25].

2.2 Rumus Matematika

2.2.1 Regresi Linear

Secara matematis rumus matematika dari regresi linear dapat dituliskan seperti pada persamaan 1 [26].

$$Y = c + mX \quad \#(1)$$

Keterangan:

Y = variabel akibat

X = variabel penyebab

m = konstanta

c = koefisien regresi

untuk nilai m dapat dihitung dengan menggunakan persamaan 2.

$$m = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad \#(2)$$

2.2.2 Random Forest

Secara matematis rumus matematika dari *random forest* dapat dituliskan seperti pada persamaan 3 dan persamaan 4 [27].

$$Entropy(Y) = - \sum_i p(Y) \log^2 p(Y) \quad \#(3)$$

Keterangan:

Y = Himpunan kasus

P(c|Y) = proporsi nilai Y terhadap kelas c.

$$Information\ Gain(Y, a) = Entropy(Y) - \sum_{v \in Value(a)} \frac{|Y_v|}{|Y|} Entropy(Y_v) \quad \#(4)$$

Keterangan:

Values(a) = nilai yang mungkin ada dalam himpunan kasus a

Y_v = sub kelas dari Y dengan kelas v yang berhubungan dengan kelas a

Y_a = semua nilai yang sesuai dengan a

2.2.3 eXtreme Gradient Boosting

Secara matematis rumus matematika dari *eXtreme Gradient Boosting* dapat dituliskan seperti pada persamaan 5, persamaan 6, persamaan 7, dan persamaan 8 [28].

$$\hat{y}_i = \sum_{t=1}^n f_t(x_i) f_t \in F \quad \#(5)$$

Keterangan:

$\hat{y}_i$ = nilai prediksi

X_i = jumlah sampel masukan

$f_t(x_i)$ = jumlah iterasi

Iterasi mempertahankan model asli dan menambahkan fungsi baru ke dalam model. Sebuah fungsi berhubungan dengan sebuah pohon, dan pohon yang baru dibuat sesuai dengan residu dari prediksi terakhir. Proses iterasi ditunjukkan pada rumus (6).

$$\hat{y}_i^{(0)} = 0 \quad \hat{y}_i^{(1)} = f_1(x_i) \quad \hat{y}_i^{(t)} = \hat{y}_i^{(y-i)} + f_t(x_i) \quad \#(6)$$

Fungsi objektif XGBoost ditunjukkan pada rumus (7) dan (8).

$$Obj = \sum_{i=1}^n l(y, \hat{y}) + \sum_{k=1}^K \Omega(f_k) \quad \#(7)$$

$$\Omega(f_k) = \gamma T + \lambda \frac{1}{2} \sum_{j=1}^T \omega_j^2 \quad \#(8)$$

2.2.4 R²

Secara matematis rumus matematika dari R² dapat dituliskan seperti pada persamaan 8 [25].

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (\underline{Y} - Y_i)^2} \quad \#(9)$$

Keterangan:

X = nilai prediksi

Y = nilai aktual

2.2.5 Mean absolute error (MAE)

Secara matematis rumus matematika dari MAE dapat dituliskan seperti pada persamaan 9 [25].

$$MAE = \frac{1}{m} \sum_{i=1}^m |X_i - Y_i| \quad \#(9)$$

Keterangan:

X = nilai prediksi

Y = nilai aktual

2.2.5 Root mean square error (RMSE)

Secara matematis rumus matematika dari RMSE dapat dituliskan seperti pada persamaan 10 [25].

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2} \quad \#(10)$$

Keterangan:

X = nilai prediksi

Y = nilai aktual

3. HASIL DAN PEMBAHASAN

Berdasarkan dari eksperimen yang dilakukan dengan perbandingan data latih dan data uji sebesar 70:30 dengan random_state 20 didapatkan hasil seperti tabel 1. Tabel 2 perbandingan data latih dan data uji sebesar 70:30 dengan random_state 42. Tabel 3 perbandingan data latih dan data uji sebesar 70:30 dengan random_state 52. Tabel 4 perbandingan data latih dan data uji sebesar 70:30 dengan random_state 78.

MODEL	R ²	MAE	RMSE
Linear Regression	-97180.488	63.506	63.507
XGBoost	0.0015	0.152	0.203
Random forest	-0.0513	0.172	0.208

MODEL	R ²	MAE	RMSE
Linear Regression	-20064.38	29.949	29.9504
XGBoost	0.0159	0.1578	0.209
Random forest	-0.0831	0.1859	0.22

MODEL	R ²	MAE	RMSE
Linear Regression	-15551.873	26.29	26.291
XGBoost	0.0071	0.161	0.21
Random forest	-0.0464	0.182	0.215

MODEL	R ²	MAE	RMSE
Linear Regression	-68647.1	53.7806	53.781
XGBoost	-0.0014	0.1533	0.2054
Random forest	-0.412	0.207	0.243

Berdasarkan hasil eksperimen sebanyak 10 kali dan dirata rata didapatkan hasil seperti pada tabel 5. Di mana rentang random state berada di antara 10 sampai 300.

MODEL	R ²	MAE	RMSE
Linear Regression	-3647.1	53.7806	53.781
XGBoost	0.014	0.125	0.1024
Random forest	-0.082	0.307	0.333

Berdasarkan hasil eksperimen didapatkan hasil yang paling bagus adalah XGBoost di mana R², MAE, dan RMSE yang paling bagus diantara yang lain. Hasil paling buruk didapatkan pada algoritma Linear Regression, di mana hasil R² yang paling kecil dan hasil MAE dan RMSE yang paling besar.

4. KESIMPULAN

Dengan membandingkan tiga metode—*Linear Regression*, *Random Forest*, dan XGBoost—penelitian ini memanfaatkan algoritma pembelajaran mesin untuk memprediksi harga properti di Brooklyn, New York. 20.894 rekaman penjualan properti digunakan. Eksperimen dilakukan dengan membagi data uji dan latih menjadi 70:30 dan menggunakan metrik evaluasi R², MAE, RMSE, dan waktu komputasi. Hasil menunjukkan bahwa XGBoost memiliki performa terbaik dengan nilai R² tertinggi dan eror (MAE, RMSE) terendah, sementara *Linear Regression* memiliki performa terburuk. Studi ini menegaskan bahwa pemilihan algoritma yang tepat untuk meningkatkan akurasi prediksi dalam analisis harga properti sangat penting.

DAFTAR PUSTAKA

- [1] USA Government, "County Population Totals: 2020-2023." Accessed: Dec. 03, 2024. [Online]. Available: <https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html>
- [2] USA Government, "Brooklyn borough, Kings County, New York - Census Bureau Profile." Accessed: Dec. 03, 2024. [Online]. Available: https://data.census.gov/profile/Brooklyn_borough,_Kings_County,_New_York?g=060XX00US3604710022

- [3] J. Gallin, "The Long-Run Relationship between House Prices and Income: Evidence from Local Housing Markets," *Real Estate Economics*, vol. 34, no. 3, pp. 417–438, Aug. 2006, doi: 10.1111/j.1540-6229.2006.00172.x.
- [4] G. V. Engelhardt and J. M. Poterba, "House prices and demographic change: Canadian evidence," *Reg Sci Urban Econ*, vol. 21, no. 4, pp. 539–546, Dec. 1991, doi: 10.1016/0166-0462(91)90017-H.
- [5] D. Gale, "Mathematica Scandinavica THE LAW OF SUPPLY AND DEMAND," 1955. [Online]. Available: <https://www.jstor.org/stable/24490348>
- [6] Q. Bi, K. E. Goodman, J. Kaminsky, and J. Lessler, "What is Machine Learning? A Primer for the Epidemiologist," *Am J Epidemiol*, Oct. 2019, doi: 10.1093/aje/kwz189.
- [7] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, "Machine learning on big data: Opportunities and challenges," *Neurocomputing*, vol. 237, pp. 350–361, May 2017, doi: 10.1016/j.neucom.2017.01.026.
- [8] B. Zheng, S. Yoon, and S. S. Lam, "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms," *Expert Syst Appl*, vol. 41, no. 4, pp. 1476–1482, Mar. 2014, doi: 10.1016/j.eswa.2013.08.044.
- [9] Stacyana Jesika, Suci Ramadhani, and Yohanna Permata Putri, "Implementasi Model Machine Learning dalam Mengklasifikasi Kualitas Air," *Jurnal Ilmiah Dan Karya Mahasiswa*, vol. 1, no. 6, pp. 382–396, Nov. 2023, doi: 10.54066/jikma.v1i6.1162.
- [10] E. Kavlakoglu and E. Russi, "What is XGBoost?," IBM.
- [11] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Syst Appl*, vol. 134, pp. 93–101, Nov. 2019, doi: 10.1016/J.ESWA.2019.05.028.
- [12] T. M. H. Hope, "Linear regression," *Machine Learning: Methods and Applications to Brain Disorders*, pp. 67–81, Jan. 2020, doi: 10.1016/B978-0-12-815739-8.00004-3.
- [13] H. I. Okagbue, P. I. Adamu, S. A. Bishop, E. C. M. Obasi, and A. O. Akinola, "Curve estimation models for estimation and prediction of impact factor and citescore using the journal percentiles: A case study of telecommunication journals," *International journal of online and biomedical engineering*, vol. 15, no. 14, pp. 31–41, 2019, doi: 10.3991/ijoe.v15i14.11373.
- [14] J. Qi, J. Du, S. M. Siniscalchi, X. Ma, and C.-H. Lee, "On Mean Absolute Error for Deep Neural Network Based Vector-to-Vector Regression," *IEEE Signal Process Lett*, vol. 27, pp. 1485–1489, 2020, doi: 10.1109/LSP.2020.3016837.
- [15] D. S. K. Karunasingha, "Root mean square error or mean absolute error? Use their ratio as well," *Inf Sci (N Y)*, vol. 585, pp. 609–629, Mar. 2022, doi: 10.1016/j.ins.2021.11.036.
- [16] P. K. Ozili, "The acceptable R-square in empirical modelling for social science research," 2023.
- [17] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geosci Model Dev*, vol. 15, no. 14, pp. 5481–5487, Jul. 2022, doi: 10.5194/gmd-15-5481-2022.
- [18] W. Darmalaksana, "Metode Penelitian Kualitatif," *Pre-Print Digital Library UIN Sunan Gunung Djati Bandung*, 2020.
- [19] H. Syahrizal and M. S. Jailani, "Jenis-Jenis Penelitian Dalam Penelitian Kuantitatif dan Kualitatif," *Jurnal QOSIM Jurnal Pendidikan Sosial & Humaniora*, vol. 1, no. 1, pp. 13–23, May 2023, doi: 10.61104/jq.v1i1.49.
- [20] R. T. Mauli, C. Simorangkir, and M. Maulista, "TUGAS BESAR 2 STATISTIK ANALISIS DATA TIME SERIES KELOMPOK 2," 2023. [Online]. Available: <https://www.researchgate.net/publication/375792430>
- [21] I. Murni, A. S. Br pa, B. R. Lubis, and A. Ikhwan, "Pengamanan Pesan Rahasia dengan Algoritma Vigenere Cipher Menggunakan PHP," *Journal on Education*, vol. 5, no. 2, pp. 3466–3476, Jan. 2023, doi: 10.31004/joe.v5i2.1027.
- [22] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, "Linear Regression," 2023, pp. 69–134. doi: 10.1007/978-3-031-38747-0_3.
- [23] R. Genuer and J.-M. Poggi, *Random Forests with R*. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-56485-8.
- [24] C. Li et al., "Power Load Forecasting Based on the Combined Model of LSTM and XGBoost," in *Proceedings of the 2019 the International Conference on Pattern Recognition and Artificial Intelligence*, New York, NY, USA: ACM, Aug. 2019, pp. 46–51. doi: 10.1145/3357777.3357792.
- [25] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput Sci*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.
- [26] T. N. Putri, A. Yordan, and D. H. Lamkaruna, "Peramalan Penerimaan Mahasiswa Baru Universitas Samudra Menggunakan Metode Regresi Linear Sederhana," 2019. [Online]. Available: <https://data.unsam.ac.id/?op=pmb>,

- [27] G. Arther Sandag, "Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest Application Rating Prediction on App Store using Random Forest Algorithm," *Cogito Smart Journal* |, vol. 6, no. 2, 2020, [Online]. Available: <https://www.kaggle.com/>
- [28] M. Dong, L. Yao, X. Wang, B. Benatallah, S. Zhang, and Q. Z. Sheng, "Gradient Boosted Neural Decision Forest," *IEEE Trans Serv Comput*, vol. 16, no. 1, pp. 330–342, Jan. 2023, doi: 10.1109/TSC.2021.3133673.