

## ANALISA PREDIKSI HARGA JUAL RUMAH MENGGUNAKAN ALGORITMA RANDOM FOREST MACHINE LEARNING

Warjiyono<sup>1)</sup>, Amin Nur Rais<sup>2)</sup>, Ibnu Alfarobi<sup>3)</sup>, Sofian Wira Hadi<sup>4)</sup>, Wawan Kurniawan<sup>5)</sup>

<sup>1)</sup> Sistem Informasi Akuntansi Kampus Kota Tegal, <sup>2)</sup> Teknologi Komputer, <sup>3)</sup> Ilmu Komputer, <sup>4)</sup> Sistem Informasi  
<sup>1,2,3,4,5)</sup> Universitas Bina Sarana Informatika

Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, DKI Jakarta

e-mail: [warjiyono.wrj@bsi.ac.id](mailto:warjiyono.wrj@bsi.ac.id)<sup>1)</sup>, [amin.arv@bsi.ac.id](mailto:amin.arv@bsi.ac.id)<sup>2)</sup>, [ibnu.iba@bsi.ac.id](mailto:ibnu.iba@bsi.ac.id)<sup>3)</sup>, [Sofian.sod@bsi.ac.id](mailto:Sofian.sod@bsi.ac.id)<sup>4)</sup>, [wawan.wwk@bsi.ac.id](mailto:wawan.wwk@bsi.ac.id)<sup>5)</sup>

\* Korespondensi: e-mail: [warji.wrj@bsi.ac.id](mailto:warji.wrj@bsi.ac.id)

### ABSTRAK

Kurangnya pengetahuan dan informasi mengenai harga rumah membuat penjual dan pembeli kesulitan dalam menentukan harga rumah. Harga rumah ditentukan oleh banyak faktor, kenaikan harga rumah pada tahun depan dan seterusnya tidak bisa diprediksi sehingga perlu dibuatnya analisis serta prediksi harga rumah. Berdasarkan permasalahan tersebut maka dilakukan penelitian dengan menggunakan metode Random Forest untuk memprediksi harga rumah sehingga memudahkan transaksi penjual dan pembeli. Metode random forest membangun model dengan menggunakan beberapa pohon keputusan secara acak dan menggabungkan prediksi dari setiap pohon untuk mendapatkan hasil prediksi. Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari website Kagel. Hasil penelitian ini menunjukkan nilai akurasi sebesar 75,10%. Perlunya dirancang suatu aplikasi web yang dapat digunakan untuk memprediksi harga rumah dengan menggunakan kriteria yang diinginkan, sehingga calon penjual atau pembeli rumah dapat memperoleh informasi harga rumah yang diinginkan dengan cepat dan mudah. mempengaruhi harga rumah secara berurutan adalah Luas Tanah (LT), Luas Bangunan (LB), Jumlah Kamar Tidur (JKT), Jumlah Kamra Mandi (JKM) dan Garasi. Hasil prediksi menunjukan adanya penurunan harga rata-rata penurunan harga 16%.

**Kata Kunci:** Harga Rumah, Algoritma Random Forest, Machine Learning

### ABSTRACT

*Lack of knowledge and information regarding house prices makes it difficult for sellers and buyers to determine house prices. House prices are determined by many factors, the increase in house prices next year and beyond cannot be predicted so it is necessary to make house price analyzes and predictions. Based on these problems, research was carried out using the Random Forest method to predict house prices so as to facilitate seller and buyer transactions. The random forest method builds a model by using several random decision trees and combining predictions from each tree to get prediction results. The data used in this research is secondary data sourced from the Kagel website. The results of this research show an accuracy value of 75.10%. It is necessary to design a web application that can be used to predict house prices using the desired criteria, so that prospective home sellers or buyers can obtain the desired house price information quickly and easily. The factors that influence house prices are land area (LT), building area (LB), number of bedrooms (JKT), number of bathrooms (JKM) and garage. The prediction results show an average price decline of 16%..*

**Keywords:** House Prices, Random Forest Algorithm, Machine Learning.

## I. PENDAHULUAN

Seiring dengan perkembangan jaman dan perkembangan teknologi, sektor properti sudah memanfaatkan internet untuk pemasaran penjualan rumah baik berbasis website maupun mobile. Saat ini banyak sekali aplikasi yang memberikan penawaran jual rumah, seperti aplikasi lamudi.co.id, rumah123.com, rumah.trovit.co.id dan masih banyak lagi.

Rumah merupakan bagian dari kebutuhan utama bagi manusia baik individu maupun yang sudah berkeluarga. Dari tahun ke tahun harga rumah dipastikan akan terus naik didorong oleh pertumbuhan jumlah penduduk, kenaikan permintaan rumah dimasyarakat, ketersediaan tanah serta kenaikan harga bangunan. Namun kenaikan harga rumah pada tahun depan dan seterusnya tidak bisa diprediksi sehingga perlu dibuatnya analisis serta prediksi [1] untuk memperkirakan harga jual rumah yang akurat agar menguntungkan [2].

Berdasarkan penelitian sebelumnya, prediksi harga rumah yang akurat merupakan tantangan besar [3]. Komponen kompleks prediksi harga rumah ditentukan oleh luas tanah, umur properti, jarak ke pusat kota, fasilitas, kualitas properti, lokasi [4] [5], aksesibilitas lokasi, struktur fisik properti dan waktu [5], luas bangunan, jumlah kamar tidur, jumlah kamar mandi, dan garasi [6] [7].

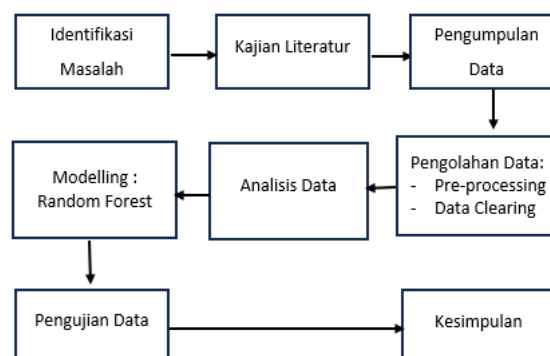
Untuk memudahkan memprediksi harga jual rumah, maka diperlukan suatu model prediksi yang dapat digunakan untuk mengetahui kisaran harga rumah yang diinginkan pembeli sesuai dengan kriteria tipe rumah yang diinginkannya. Sedangkan bagi penjual, model prediksi dapat membantu penjual rumah yang kesulitan menentukan harga rumah berdasarkan harga pasar.

Salah satu metode yang dapat digunakan untuk memprediksi harga rumah adalah metode random forest machine learning. Kualitas penilaian indeks harga rumah dapat ditingkatkan secara signifikan dengan menggunakan metode random forest [8]. Machine learning banyak dipakai dalam penelitian keuangan untuk memperkirakan harga rumah [9] Metode random forest merupakan bagian dari metode data mining yang metode dasarnya adalah pohon keputusan. Random forest membangun model dengan menggunakan beberapa pohon keputusan secara acak dan menggabungkan prediksi dari setiap pohon untuk mendapatkan hasil prediksi.

Tujuan penelitian ini adalah untuk mengetahui hasil prediksi harga rumah di wilayah Jakarta Selatan dengan metode Random Forest Regression. Penelitian terdahulu dengan data set yang sama telah dilakukan namun menggunakan metode Multiple Linier Regresion [1]. Manfaat memberikan informasi kisaran harga bagi penjual dan pembeli saat melakukan transaksi jual beli rumah.

## II. METODOLOGI PENELITIAN

Penelitian ini menggunakan data sekunder yang berasal dari website penyedia data dengan alamat <https://www.kaggle.com/datasets/wisnuanggara/register-harga-rumah>. Populasi yang digunakan dalam penelitian ini adalah seluruh harga rumah di Jakarta Selatan. Sedangkan sampel yang digunakan disini adalah harga rumah di wilayah Jakarta Selatan. Data yang digunakan adalah 1001 data penjualan rumah di Jakarta Selatan yang diakses pada tanggal 18 September 2023.



Gambar 1. Tahapan Penelitian

Gambar 1 merupakan tahapan penelitian yang dimulai dari identifikasi masalah, dilanjutkan dengan kajian literatur yaitu merujuk pada jurnal-jurnal internasional bereputasi dan beberapa jurnal nasional. Selanjutnya mencari data publik terkait harga rumah yang dinabil dari website kaggle. Data tersebut diolah untuk tahap pre-processing dan data clearing untuk menghilangkan data yang tidak relevan atau data tidak lengkap dan amenjadi data yang lebih baik. Setelah itu dilakukan analisis deskriptif dan dilanjutkan untuk dibuat modelling menggunakan algoritma random forest. Tahap selanjutnya dilakukan pengujian dan kesimpulan hasil.

### III. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan aplikasi berbasis website yaitu Google Colab untuk melakukan data cleaning, transformasi data, visualisasi data, machine learning, alat simulasi.

#### A. Pre-Processing

Data yang diperoleh untuk pelatihan dan pengujian model harus dijelaskan dengan tepat sebelum membuat model sehingga model dapat mempelajari polanya dengan lebih cepat [4]. Tahap pre-procesing akan membuat dataset akhir, yang akan dimasukkan ke dalam model yang dibuat [10]. Data yang diolah dalam penelitian ini sebanyak 1001 data dan 7 (tujuh) kolom yaitu Harga, Luas Tanah (LT), Luas Bangunan (LB), Jumlah Kamar Tidur (JKT), Jumlah Kamar Mandi (JKM), Garasi (GRS) dan Kota.

	HARGA	LT	LB	JKT	JKM	GRS	KOTA
0	28000000000	1100	700	5	6	ADA	JAKSEL
1	19000000000	824	800	4	4	ADA	JAKSEL
2	4700000000	500	400	4	3	ADA	JAKSEL
3	4900000000	251	300	5	4	ADA	JAKSEL
4	28000000000	1340	575	4	5	ADA	JAKSEL
...	...	...	...	...	...	...	...
996	16000000000	488	550	6	5	ADA	JAKSEL
997	4500000000	209	270	4	4	ADA	JAKSEL
998	29000000000	692	400	4	3	TIDAK ADA	JAKSEL
999	1700000000	102	140	4	3	TIDAK ADA	JAKSEL
1000	1250000000	63	110	3	3	TIDAK ADA	JAKSEL

1001 rows × 7 columns

Gambar 2. Pre-Processing

#### B. Data Cleaning

Selanjutnya dilakukan data cleaning yang bertujuan untuk menghapus data-data yang tidak relevan. Dari 1001 data yang ada ditemukan data duplicate, maka dilakukan proses data cleaning terlebih dahulu agar tidak ada data yang sama atau duplikat. Tabel 2 menunjukkan 931 data relevan setelah dilakukan penghapusan terhadap data duplikat.

	HARGA	LT	LB	JKT	JKM	GRS	KOTA
0	28000000000	1100	700	5	6	ADA	JAKSEL
1	19000000000	824	800	4	4	ADA	JAKSEL
2	4700000000	500	400	4	3	ADA	JAKSEL
3	4900000000	251	300	5	4	ADA	JAKSEL
4	28000000000	1340	575	4	5	ADA	JAKSEL
...	...	...	...	...	...	...	...
996	16000000000	488	550	6	5	ADA	JAKSEL
997	4500000000	209	270	4	4	ADA	JAKSEL
998	29000000000	692	400	4	3	TIDAK ADA	JAKSEL
999	1700000000	102	140	4	3	TIDAK ADA	JAKSEL
1000	1250000000	63	110	3	3	TIDAK ADA	JAKSEL

931 rows × 7 columns

Gambar 3. Data Hasil Cleaning

### C. Analisis Deskriptif

Penelitian ini melakukan analisis deskriptif pada kumpulan data, dan hasil analisis ditunjukkan pada Tabel 3. Pada tabel 3 menunjukkan dari jumlah data sebesar 931 dapat dilihat bahwa rata-rata harga total rumah di Jakarta Selatan adalah 1,75 milyar rupiah dengan luas tanah rata-rata adalah 535 m2 dan luas bangunan 488 m2, jumlah kamar tidur 4 dan jumlah kamar mandi 3.

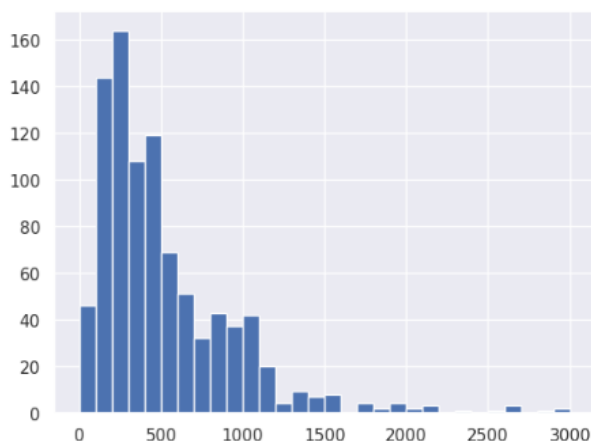
	HARGA	LT	LB	JKT	JKM
count	9.310000e+02	931.000000	931.000000	931.000000	931.000000
mean	1.754738e+10	535.464017	488.788400	4.47261	3.921590
std	2.120253e+10	543.800947	465.069974	2.06300	1.943635
min	4.300000e+08	22.000000	38.000000	1.00000	1.000000
25%	6.750000e+09	217.000000	300.000000	4.00000	3.000000
50%	1.300000e+10	400.000000	400.000000	4.00000	4.000000
75%	2.000000e+10	682.500000	600.000000	5.00000	4.000000
max	2.500000e+11	6790.000000	10000.000000	27.00000	27.000000

Gambar 4. Analisis deskriptif

### D. Analisis Visual

Dalam penelitian ini harga rumah digunakan sebagai variabel dependen. Sedangkan variabel independen adalah Luas Tanah (LT), Luas Bangunan (LB), Jumlah Kamar Tidur (JKT), Jumlah Kamar Mandi (JKM), Garasi (GRS) dan Kota. Kami mempelajari pengaruh variabel independen terhadap variabel dependen

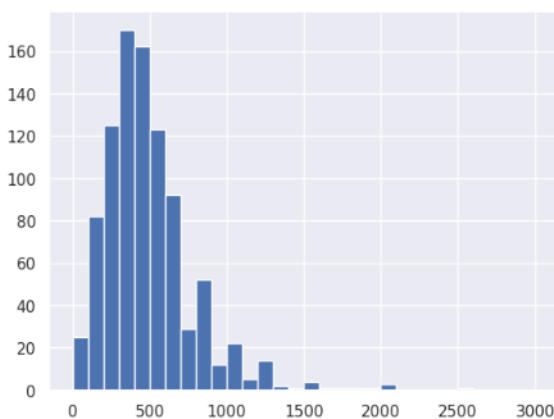
#### 1. Jumlah Luas Tanah



Gambar 5. Jumlah Luas Tanah

Gambar 2. menunjukkan bahwa luas tanah 300 m2 paling banyak yang dijual di Jakarta Selatan.

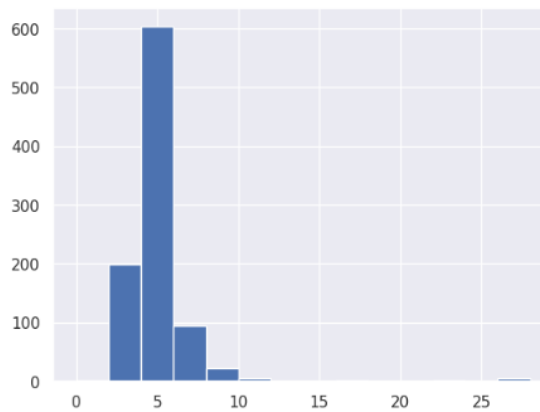
#### 2. Jumlah Luas Bangunan



Gambar 6. Jumlah Luas Bangunan

Gambar 3. menunjukan bahwa luas bangunan 400 m2 paling banyak yang dijual di Jakarta Selatan.

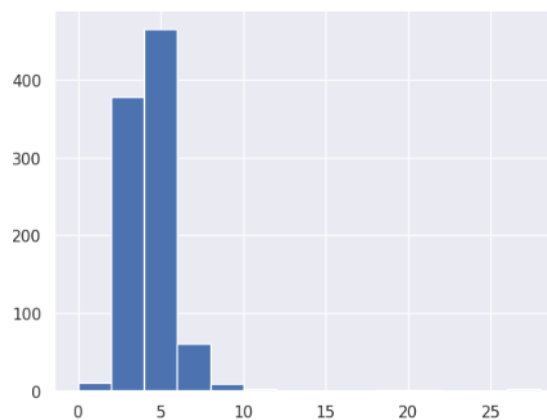
### 3. Jumlah Kamar Tidur



Gambar 7. Jumlah Kamar Tidur

Gambar 4. menunjukkan bahwa jumlah 5 kamar tidur paling banyak yang dijual di Jakarta Selatan.

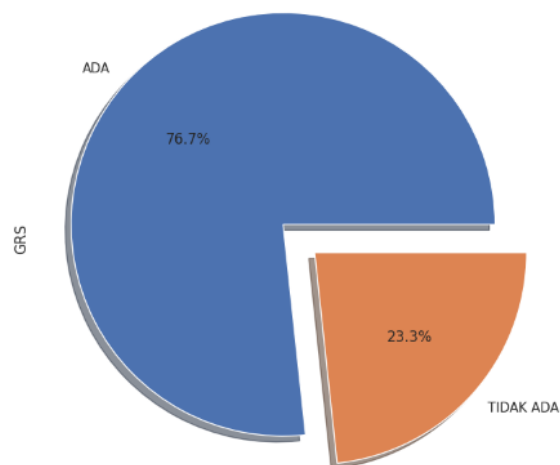
### 4. Jumlah kamar Mandi



Gambar 8. Jumlah Kamar Mandi

Gambar 5. menunjukkan bahwa jumlah 5 kamar mandi paling banyak yang dijual di Jakarta Selatan

### 5. Garasi Terhadap Harga Rumah



Gambar 9. Jumlah Rumah Bergarasi

Gambar 9. menunjukkan jumlah rumah di Jakarta Selatan yang mempunyai garasi sebanyak 76,7% sedangkan rumah yang tidak punya garasi sebanyak 23,3%.

## 6. Faktor Yang Berpengaruh



Gambar 10. Faktor Yang Berpengaruh

Gambar 7. menunjukkan bahwa yang mempengaruhi harga rumah secara berurutan adalah Luas Tanah (LT), Luas Bangunan (LB), Jumlah Kamar Tidur (JKT), Jumlah Kamra Mandi (JKM) dan Garasi.

## 7. Akurasi

```
[31] score = accuracy_score(test_y ,RF_predictions)
      print('Accuracy Random Forest Model:',score)

Accuracy Random Forest Model: 0.7510729613733905
```

Gambar 11. Hasil pengujian

Dari pengujian diatas pada gambar 8 terlihat bahwa Akurasi yang didapat sebesar 0.07510 atau 75,10%, cukup baik untuk memprediksikan harga rumah berdasarkan variable independent yang ada.

## E. Pengujian Model

Setelah melatih model dengan pelatihan dataset, tahap penelitian selanjutnya adalah menguji kecakapan model prediktif. Hal ini dicapai dengan menghapus harga sebenarnya dari kumpulan data dan memperkirakan model untuk memprediksi harga rumah. Harga rumah yang diprediksi dan sebenarnya kemudian digabungkan dan selisihnya dihitung. Hal ini terlihat pada Tabel 4. Hasil yang diperoleh menunjukkan ada perbedaan antara harga aktual dan harga prediktif, dalam kolom selisih terlihat ada penurunan harga yang sangat signifikan, rata-rata penurunan harga 16%.

	Aktual	Prediksi	Selisih
408	6000000000	1.132716e+10	5.327160e+09
525	12000000000	1.825589e+10	6.255892e+09
948	28000000000	2.377689e+10	-4.223110e+09
980	25000000000	1.877013e+10	-6.229868e+09
506	20000000000	1.767919e+10	-2.320810e+09
...	...	...	...
802	57000000000	4.152200e+10	-1.547800e+10
470	42000000000	2.310402e+09	-1.889598e+09
592	65000000000	6.492509e+09	-7.490714e+06
118	35000000000	3.012035e+09	-4.879650e+08
76	18000000000	1.762964e+10	-3.703600e+08

280 rows x 3 columns

Gambar 12. Hasil Pengujian Aktual vs Prediksi

#### IV. KESIMPULAN

Kurangnya pengetahuan dan informasi mengenai harga rumah membuat penjual dan pembeli kesulitan dalam menentukan harga rumah. Harga rumah ditentukan oleh banyak faktor, kenaikan harga rumah pada tahun depan dan seterusnya tidak bisa diprediksi sehingga perlu dibuatnya analisis serta prediksi harga rumah. Berdasarkan permasalahan tersebut maka dilakukan penelitian dengan menggunakan metode Random Forest untuk memprediksi harga rumah sehingga memudahkan transaksi penjual dan pembeli. Metode random forest membangun model dengan menggunakan beberapa pohon keputusan secara acak dan menggabungkan prediksi dari setiap pohon untuk mendapatkan hasil prediksi. Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari website Kagel. Hasil penelitian ini menunjukkan nilai akurasi sebesar 75,10%. Perlunya dirancang suatu aplikasi web yang dapat digunakan untuk memprediksi harga rumah dengan menggunakan kriteria yang diinginkan, sehingga calon penjual atau pembeli rumah dapat memperoleh informasi harga rumah yang diinginkan dengan cepat dan mudah. mempengaruhi harga rumah secara berurutan adalah Luas Tanah (LT), Luas Bangunan (LB), Jumlah Kamar Tidur (JKT), Jumlah Kamar Mandi (JKM) dan Garasi. Hasil prediksi menunjukkan adanya penurunan harga rata-rata penurunan harga 16%.

#### DAFTAR PUSTAKA

- [1] M. L. Mu'tashim, T. Muhayat, S. A. Damayanti, H. N. Zaki, dan R. Wirawan, "Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression," *Inform. J. Ilmu Komput.*, vol. 17, no. 3, hal. 238, 2021, doi: 10.52958/iftk.v17i3.3635.
- [2] S. Lahmiri, S. Bekiros, dan C. Avdoulas, "A comparative assessment of machine learning methods for predicting housing prices using Bayesian optimization," *Decis. Anal. J.*, vol. 6, no. November 2022, hal. 100166, 2023, doi: 10.1016/j.dajour.2023.100166.
- [3] C. Zhan, Y. Liu, Z. Wu, M. Zhao, dan T. W. S. Chow, "A hybrid machine learning framework for forecasting house price," *Expert Syst. Appl.*, vol. 233, no. July, hal. 120981, 2023, doi: 10.1016/j.eswa.2023.120981.
- [4] A. B. Adetunji, O. N. Akande, F. A. Ajala, O. Oyewo, Y. F. Akande, dan G. Oluwadara, "House Price Prediction using Random Forest Machine Learning Technique," *Procedia Comput. Sci.*, vol. 199, hal. 806–813, 2021, doi: 10.1016/j.procs.2022.01.100.
- [5] D. Rey-Blanco, J. L. Zofío, dan J. González-Arias, "Improving hedonic housing price models by integrating optimal accessibility indices into regression and random forest analyses," *Expert Syst. Appl.*, vol. 235, no. August 2023, hal. 121059, 2024, doi: 10.1016/j.eswa.2023.121059.
- [6] A. Saiful, "Prediksi Harga Rumah Menggunakan Web Scrapping dan Machine Learning Dengan Algoritma Linear Regression," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 1, hal. 41–50, 2021, doi: 10.35957/jatisi.v8i1.701.
- [7] E. S. Lestari dan I. Astuti, "Penerapan Random Forest Regression Untuk Memprediksi Harga Jual Rumah Dan Cosine Similarity Untuk Rekomendasi Rumah Pada Provinsi Jawa Barat," *J. Ilm. FIFO*, vol. 14, no. 2, hal. 131, 2022, doi: 10.22441/fifo.2022.v14i2.003.
- [8] J. Hong, H. Choi, dan W. S. Kim, "A house price valuation based on the random forest approach: The mass appraisal of residential property in south korea," *Int. J. Strateg. Prop. Manag.*, vol. 24, no. 3, hal. 140–152, 2020, doi: 10.3846/ijspm.2020.11544.
- [9] Y. Zheng, "Community resilience and house prices: A machine learning approach," *Financ. Res. Lett.*, vol. 58, no. PB, hal. 104400, 2023, doi: 10.1016/j.frl.2023.104400.
- [10] I. Kurniawan, D. C. P. Buani, A. Abdussomad, W. Apriliah, dan R. A. Saputra, "Implementasi Algoritma Random Forest Untuk Menentukan Penerima Bantuan Raskin," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 2, hal. 421–428, 2023, doi: 10.25126/jtiik.20231026225.