Prediksi Harga Rumah Di Jabodetabek Menggunakan Metode Artificial Neural Network

Muhammad Abdullah Hafizh[1], Subairi, Raditya Dimas Libriawan[2], Naufal Duta Maulana[3], Agung Mustika Rizki[4]*

[1,2,3,4]Program Studi Informatika, Universitas Pembangunan Nasional Veteran Jawa Timur

** Penulis Korespondensi : 21081010099@student.upnjatim.ac.id*

ABSTRACT

A house is a fundamental need for humans. Determining house prices is a crucial aspect of property transactions, especially in major areas like Jabodetabek, where property prices are consistently rising. Prediction is a suitable tool to assist in decision-making for determining house prices. There are numerous methods that can be applied for prediction; the author employs the Artificial Neural Network (ANN) method. ANN is known as a highly flexible predictive algorithm capable of accommodating various input features. The results of using the ANN method in predicting house prices in the Jabodetabek show a small difference between the actual values and the predicted values, with a Mean Absolute Error (MAE) of 0.209, a Mean Squared Error (MSE) of 0.159, and a Mean Absolute Percentage Error (MAPE) of 4.951. The small differences indicate the ANN method is effective in predicting house prices in the Jabodetabek.

Article History

Received: 08-12-2024
Revised : 19-12-2024
Accepted: 30-12-2024

Keywords

Prediksi
Harga Rumah
Spesifikasi
Artifical Neural Network
Jabodetabek

ABSTRAK

Rumah adalah kebutuhan pokok bagi manusia. Penentuan harga rumah tentu menjadi salah satu aspek yang krusial dalam transaksi jual beli properti, terutama di wilayah besar seperti Jabodetabek dengan harga properti yang selalu meningkat. Prediksi adalah salah satu alat bantu yang sesuai untuk alat bantu pemberi keputusan dalam penentuan harga rumah. Terdapat banyak sekali metode yang dapat diterapkan dalam prediksi, Pada penelitian ini menggunakan metode *Artificial Neural Network (*ANN). Algoritma ANN dikenal sebagai salah satu metode prediksi yang cukup fleksibel dengan berbagai *input* fitur. Hasil dari penggunaan metode ANN dalam prediksi harga rumah di wilayah Jabodetabek memperlihatkan selisih antara nilai asli dan nilai prediksi yang kecil yaitu *Mean Absolute Erro*r (MAE) sebesar 0.209, nilai *Mean Squared Error* (MSE) sebesar 0.159, dan nilai *Mean Absolute Percentage Error* (MAPE) sebesar 4.951. Dari selisih nilai yang kecil meunjukkan bahwa metode ANN baik dalam memprediksi prediksi harga rumah di wilayah Jabodetabek.

PENDAHULUAN

Rumah adalah kebutuhan pokok bagi setiap makhluk hidup untuk tinggal dan menetap. Selain sebagai tempat tinggal, rumah juga memiliki nilai tersendiri, mencerminkan ukuran, kondisi, dan status ekonomi pemiliknya [1]. Di kota besar, seperti Jabodetabek harga properti terus meningkat karena tingginya permintaan, fasilitas umum yang memadai, serta nilai investasi jangka panjang yang menjanjikan. Dengan faktor-faktor seperti harga tanah dan rumah yang terus naik setiap tahun, investasi properti di Jabodetabek menjadi pilihan menarik bagi berbagai kalangan.

Wilayah Jabodetabek (Jakarta, Bogor, Depok, Tangerang, dan Bekasi) merupakan salah satu kawasan metropolitan terbesar di Indonesia yang mengalami pertumbuhan ekonomi dan demografi pesat, menarik banyak orang dari berbagai daerah untuk mencari peluang kerja dan meningkatkan kualitas hidup. Permintaan hunian yang tinggi dipicu oleh pertumbuhan penduduk, urbanisasi, peningkatan standar hidup, dan perkembangan infrastruktur, seperti jalan tol dan kereta api cepat, yang mempermudah aksesibilitas dan meningkatkan nilai properti. Selain itu, fasilitas umum seperti pusat perbelanjaan, rumah sakit, dan sekolah juga berkontribusi pada kenaikan harga rumah di wilayah ini [2].

Penentuan harga rumah yang akurat menjadi aspek krusial dalam transaksi jual beli properti. Harga yang tepat tidak hanya memberikan keadilan bagi penjual dan pembeli, tetapi juga

mencerminkan kondisi pasar yang sesungguhnya. Namun, penentuan harga rumah tidaklah sederhana karena dipengaruhi oleh berbagai variabel yang kompleks dan saling terkait. Beberapa faktor utama yang mempengaruhi harga rumah meliputi lokasi geografis, fasilitas yang tersedia, jumlah kamar tidur dan kamar mandi, ukuran tanah dan bangunan, jenis sertifikat, dan fasilitas pendukung seperti *carport,* dan garasi, serta kondisi properti [1], [3].

Prediksi menjadi alat bantu yang penting, kini sering dijadikan sebagai salah satu pertimbangan dalam pengambilan keputusan pembelian. Kemampuan untuk melakukan prediksi dengan tingkat akurasi yang terpercaya dapat mendukung dalam membuat keputusan yang lebih tepat [4]. Dalam konteks prediksi harga rumah, selain variabel yang mempengaruhi harga dan interaksi antar variabel tersebut, kualitas dan kelengkapan data juga menjadi faktor penting untuk memastikan akurasi model prediksi.

Penelitian terkait yang dilakukan oleh [4] tentang prediksi harga rumah dengan metode *Multiple Linear Regression* yang merupakan salah satu metode statistika. Pada penelitiannya atribut yang digunakan yaitu harga rumah, luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi, dan jumlah garasi. Hasil dari penelitian didapat nilai MAE sebesar 1980.35 dan akurasi model sebesar 0.77 atau 77%. Dari hasil tersebut, terdapat selisih yang besar antara nilai asli dan nilai prediksi, hal ini menandakan bahwa model statistik yang digunakan kurang baik dalam memprediksi harga rumah.

Oleh karena itu, penelitian ini menggunakan metode *machine learning* karena lebih baik dibandingkan metode statistik seperti *multiple linear regression* [5]. *Machine learning* berkemampuan untuk memprediksi data baru dengan mencari pola dalam data yang besar dan menarik kesimpulan berdasarkan prediksi dari data tersebut [6]. Salah satu metode *machine learning* yang dapat digunakan untuk memprediksi harga rumah dengan data yang besar yaitu *Artificial Neural Network* (ANN). ANN adalah salah satu metode yang banyak digunakan dalam analisis ini karena kemampuannya untuk menangkap hubungan linier antara satu variabel dependen yakni harga rumah dengan beberapa variabel independen seperti luas bangunan, ukuran tanah, dan jumlah ruangan [7]. Hubungan antara variabel tersebut akan menghasilkan tingkat akurasi yang disebut dengan korelasi [8]. Metode ini tidak hanya mengidentifikasi faktor yang paling berpengaruh terhadap harga rumah, tetapi juga menyediakan dasar untuk membuat prediksi harga yang akurat berdasarkan data spesifikasi rumah yang tersedia.

TINJAUAN PUSTAKA

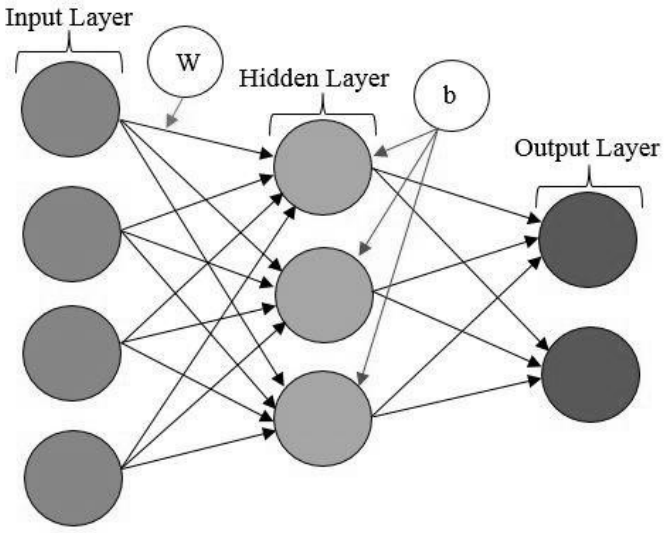
Rumah

Pada dasarnya, kebutuhan pokok manusia terbagi menjadi tiga macam, yaitu kebutuhan sandang, pangan, dan papan. Ketiga kebutuhan pokok ini harus terpenuhi dan tidak bisa diabaikan satu dengan yang lainnya. Rumah merupakan salah satu contoh kebutuhan pokok atau primer manusia pada kebutuhan papan. Secara umum, rumah dapat diartikan sebagai bangunan untuk tempat berlindung, serta tempat tumbuh dan berkembangnya manusia. Banyak faktor yang mempengaruhi setiap orang untuk memilih rumah tinggal diantaranya adalah harga rumah, luas tanah, luas bangunan, model rumah , jarak rumah ke jalan raya, jarak rumah ke tempat kerja, jarak rumah ke sekolah anak, adanya fasilitas keamanan lingkungan, jarak rumah dengan tempat belanja dan lain sebagainya [9].

Artificial Neural Network

Artificial Neural Network (ANN) merupakan algoritma *feed-forward* atau bergerak maju. Algoritma ini mampu mengenali pola dari data *input* yang berjumlah besar [10]. ANN berbasis jaringan saraf dengan penyusun yang disebut *neuron* dan penghubung antar *neuron* disebut *node*. *Node* menghubungkan antar neuron dengan menambahkan bobot yang akan mempengaruhi keakurasian model. Jaringan ANN terdiri dari tiga lapisan, yaitu *input layer*, *hidden layer*, dan *output layer*. ANN memperoleh atribut kombinasi linear dan mengubah keluaran menjadi *non-linear* dari atribut tersebut [11].

ISSN: 2774-4345, DOI: 10.31284/j.kernel.2024.v5i2.6806



Gambar 1. Arsitektur algoritma ANN
Sumber: (Ovarian Cancer Classification Accuracy Analysis Using 15-Neuron Artificial Neural Networks Model", 2019)

Pada Gambar 1, untuk menghitung bobot (W), fungsi *sigmoid* bekerja pada *input* dan bobot yang sesuai. Nilai bobot disesuaikan pada saat proses pelatihan untuk mengenali berdasarkan karakteristik dari *input*. Hubungan antara bobot dengan node disebut bias (b) [12]. Ketika mencapai *output layer,* ANN menghitung nilai *error* yang akan digunakan untuk mengubah bobot di proses selanjutnya yang disebut proses *backpropagation.* Proses *backpropagation* berjalan mundur dari *output layer* hingga *input layer* proses ini dihitung 1 iterasi. Proses selanjutnya menjalankan kembali proses *feed-forward* hingga mencapai hasil yang ditentukan.

Harga Rumah

Harga rumah harus ditentukan dengan sangat cermat dan teliti, baik dari sisi penjual ataupun pembeli yang sedang memperkirakan besaran uang yang dikeluarkan untuk sebuah rumah. Dalam menentukan harga rumah, terdapat beberapa faktor yang dapat mempengaruhi harga dari sebuah rumah tersebut, antara lain: lokasi, aksesibilitas menuju rumah, kondisi fisik, fasilitas, luas tanah, dan bangunan, usia bangunan, hingga kualitas material dan desain dari rumah tersebut. Selain itu, nilai pajak rumah, yang dikenal sebagai Nilai Jual Objek Pajak (NJOP), bervariasi di setiap daerah. Hal ini menjadikan nilai pajak sebagai faktor penting dalam proses pembelian rumah. Selain itu, NJOP yang tinggi juga dapat menjadi aset investasi bagi pemilik rumah jika mereka berencana untuk menjualnya di masa depan [13].

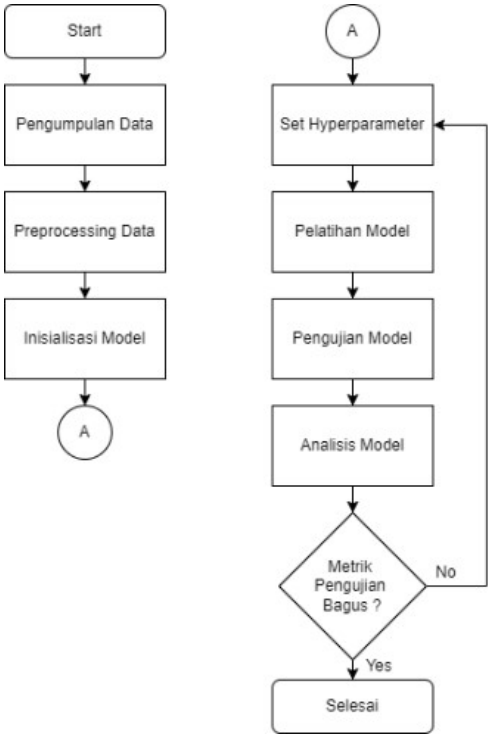
Prediksi

Prediksi adalah proses memperkirakan secara sistematis sesuatu yang akan terjadi di masa depan berdasarkan informasi dari masa lalu dan saat ini, sehingga menghasilkan estimasi yang mendekati hasil sebenarnya [14]. Data yang dapat digunakan dalam prediksi adalah data kuantitatif. Adapun prediksi ini bertujuan untuk memperkecil kesalahan yang terjadi antara perkiraan dengan kejadian yang sebenarnya.

METODE

Penelitian dilakukan dengan beberapa tahapan, yaitu pengumpulan data, *preprocessing* data, pelatihan model, pengujian model, dan analisis hasil. Untuk penjelasan lebih detail lihat pada Gambar 2.

ISSN: 2774-4345, DOI: 10.31284/j.kernel.2024.v5i2.6806



Gambar 2. Diagram alir tahapan penelitian

Pengumpulan Data

Penelitian ini menggunakan *dataset* yang diambil dari situs Kaggle (www.kaggle.com) merupakan situs yang menyediakan *dataset* publik yang dapat digunakan untuk pengembangan model. *Dataset* yang digunakan berjudul "Daftar Harga Rumah Jabodetabek". Data ini discraping dari situs rumah123.com pada akhir tahun 2022. *Dataset* ini terdiri dari 27 fitur, namun tidak semua fitur ini digunakan untuk penelitian.

Preprocessing Data

Preprocessing dan *cleaning* data dilakukan kepada *dataset* agar lebih efektif dan bisa menampilkan informasi yang relevan untuk pelatihan model nantinya. Proses ini mencakup penghapusan kolom atau variabel yang tidak diperlukan, menghapus duplikasi data, memeriksa data yang tidak konsisten, dan encoding data, agar bisa diproses lebih lanjut.

Inisialisasi Model

Inisialisasi model meliputi pembagian *dataset* (split data) menjadi data latih dan data uji, serta menentukan jumlah *layer* dan struktur model lainnya serta fungsi aktivasi. Fungsi aktivasi ReLu memberikan *output* nol jika *input* negatif dan jika *input* positif, maka *output* akan sama dengan *input*. ReLU memiliki kemampuan untuk memecahkan masalah nonlinier. Pada saat ini fungsi aktivasi ReLu paling umum digunakan untuk digunakan membangun model. Persamaan fungsi ReLu dituliskan pada persamaan (1).

$$f(x) = \max(0, x) \quad \text{..... (1)}$$

Keterangan:
x : nilai dari *input*,
$f(x)$: hasil *output* fungsi ReLU dalam bentuk 0 dan 1.

ISSN: 2774-4345, DOI: 10.31284/j.kernel.2024.v5i2.6806

Pelatihan dan Pengujian Model

Pelatihan dan pengujian model dilakukan dengan data yang sudah dibagi. Selanjutnya menentukan *hyperparameter* yang dibutuhkan model seperti *epochs*, *batch*, *learning rate*. Tidak hanya itu, persentase *split data* mungkin juga dapat berubah tergantung dari akurasi model yang dihasilkan.

Analisis Hasil

Analisis hasil dilakukan dengan melihat Metrik pengujian berupa *Mean Absolute Error* (MAE) yang menghitung selisih antara nilai aktual dan nilai prediksi. Apabila nilai MAE melebihi batas yang ditentukan maka dilakukan pelatihan ulang model dengan konfigurasi *hyperparameter* yang berbeda.

HASIL DAN PEMBAHASAN

Preprocessing Data

Tahapan pertama dari p*reprocessing* ini ialah menghapus kolom yang memiliki banyak nilai *null*. Setelah itu melakukan *cleaning data* dengan memberikan *threshold* pada data yang memiliki kamar tidur lebih dari 13. Kemudian, pada kolom *electricity* dilakukan *cleaning* dengan menghapus data yang memiliki nilai "lainnya" dan mengubah tipedata menjadi float dari yang awalnya string. Selanjutnya, fitur *furnishing* dilakukan *encoding* disertai pemberian bobot disetiap fasilitasnya sehingga luarannya ialah total dari keseluruhan bobot. Pada tahap terakhir *preprocessing* yaitu melihat korelasi antara fitur dengan harga. Apabila terdapat fitur dengan nilai yang kecil, maka fitur tersebut dapat dihilangkan.

Pelatihan dan Pengujian Model

Pelatihan dan pengujian dilakukan beberapa kali dengan *hyperparameter* seperti Tabel 1.

Tabel 1. *Hyperparameter* yang digunakan pada penelitian ini

No	Name	Value
1	*Split Data*	*Train* 90% *Test* 10%
2	*Layer*	64, 32, 16, 1 64, 32, 16, 8, 1 128, 64, 32, 16, 8, 1
3	*Optimizer*	Adam
4	*Learning Rate* (LR)	0.01 dan 0.001
5	*Epochs*	60 Epochs 120 Epochs 240 Epochs 480 Epochs 960 Epochs
6	*Batch*	32 dan 64
7	*Validation Split*	10%

ISSN: 2774-4345, DOI: 10.31284/j.kernel.2024.v5i2.6806

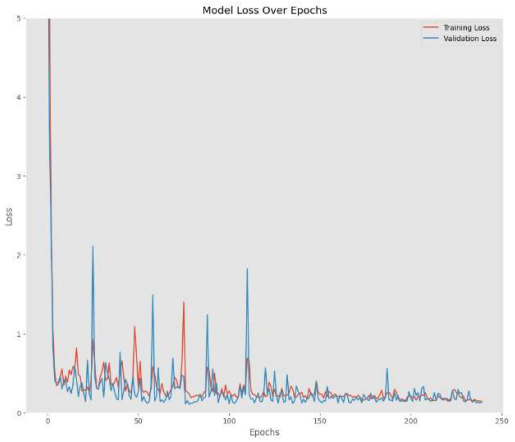
Analisis Hasil Pengujian

Analisis hasil pengujian dilakukan dengan melihat hasil MAE, MSE, dan MAPE. Apabila ketiga metrik pengujian menunjukkan hasil terkecil, maka dapat ditentukan bahwa itu merupakan model paling optimal. Pada penelitian ini model paling optimal didapat dengan epochs 240, untuk lebih detail dapat dilihat pada Tabel 2 yang menunjukkan data terbaik berdasarkan metrik pengujian MAE, MSE, dan MAPE.

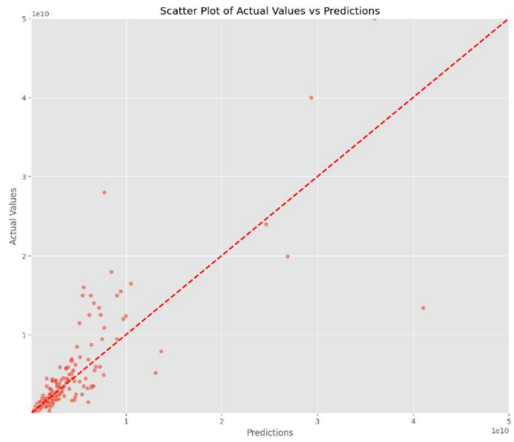
Tabel 2. Hasil pengujian dengan hasil teratas

No	*Layer*	LR	Epochs	Batch	MAE	MSE	MAPE (%)
1	64,32,16,8,1	0.01	240	32	0.375	0.230	4.932
2	128,64,32,16,8	0.01	240	32	0.209	0.159	4.951
3	64,32,16	0.01	240	32	0.300	0.181	5.118

Selanjutnya, dari ketiga data diatas dilakukan pengujian ulang dengan konfigurasi yang sama. Hasil paling optimal dan stabil ialah data nomor 2 dari Tabel 2. Dapat dilihat pada Gambar 3, antara *training loss* dan *validation loss* keduanya bergerak kebawah dan berdekatan yang menandakan bahwa model tidak terindikasi *overfitting* ataupun *underfitting* dan Gambar 4 yaitu menunjukkan persebaran data aktual dan prediksi, namun masih terdapat beberapa data yang berjauhan.



Gambar 3. Grafik *model loss*



Gambar 4. Persebaran data aktual dan prediksi

KESIMPULAN

Kinerja dari metode *Artificial Neural Network* (ANN) dalam memprediksi harga rumah di wilayah Jabodetabek dinilai cukup baik. Dari beberapa pengujian yang dilakukan dengan konfigurasi yang optimal didapatkan nilai prediksi yang cukup relevan dibanding harga asli rumah dengan spesifikasi fitur rumah yang dimiliki. Konfigurasi dari model harus dicari yang paling optimal, dengan mempertimbangkan nilai-nilai dari *hyperparameter* yang digunakan seperti *Split Data, Layer, Optimizer, Learning Rate, Epochs, Batch,* dan *Validation Split.* Hasil akhir yang paling optimal didapatkan nilai *Mean Absolute Error* (MAE) sebesar 0.209, *Mean Squared Error* sebesar 0.159, dan Mean Absolute Error (MAE) Sebesar 4.951. Hasil ini menunjukan bahwa Metode *Artificial Neural Network* (ANN) dapat digunakan untuk melakukan prediksi harga rumah dengan hasil yang cukup akurat, jika data dan konfigurasi yang digunakan relevan dan optimal.

DAFTAR PUSTAKA

[1] M. Labib Mu'tashim, S. A. Damayanti, H. N. Zaki, T. Muhayat, and R. Wirawan, "Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression," vol. 3, p. 2021.

[2] N. Agustina, C. Putri, and D. B. Arianto, "Komparasi Penggunaan Information Gain Pada Machine Learning untuk Memprediksi Harga Rumah di Jabodetabek," *Jurnal Sains dan Teknologi*, vol. 5, no. 3, pp. 756–762, doi: 10.55338/saintek.v5i1.2052.

[3] F. Febyanti, "Pemodelan Faktor-Faktor yang Mempengaruhi Harga Rumah di Jabodetabek Menggunakan Metode Regresi Probit," *Jurnal Riset Statistika*, pp. 50–56, Jul. 2022, doi: 10.29313/jrs.vi.905.

[4] A. Widyastuti, "Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Metode Multiple Linear Regression," vol. 4, no. 1, pp. 30–35, 2024, [Online]. Available: http://ejurnal.unim.ac.id/index.php/submit

[5] S. A. F. Harahap, M. F. Rafliansyah, A. Septyanto, and I. Pratama, "Studi Perbandingan Algoritma Machine Learning pada Kasus Image Classification Studi Perbandingan Algoritma Statistic, Ensemble, dan Neural Network Pada Kasus Image Classification," *Jurnal Rekayasa Elektro Sriwijaya*, vol. 4, no. 2, pp. 46–51.

[6] M. Haris Husni, R. Hendi Nofanolo Harefa, Ambiyar, and M. Agni Zaus, "Perbandingan Penggunaan Statistika Dan Machine Learning Dalam Teknik Pengolahan Data," *Jurnal Sistem dan Teknologi Informasi*, vol. 6, pp. 440–450, Aug. 2024, [Online]. Available: https://journalpedia.com/1/index.php/jsti

[7] W. Wilianto, Y. Yuliana, A. Suwandhi, J. Jimmy, and J. Putra, "Penerapan AI dalam Menentukan Harga Mobil Bekas Berdasarkan Tahun Perakitan," *Jurnal Minfo Polgan*, vol. 13, no. 1, pp. 550–560, Jun. 2024, doi: 10.33395/jmp.v13i1.13728.

[8] K. Alfin Yazi Bara Mentari, A. Falih Hartono, R. Phasca Sanggam Sinaga, A. Puspita Sari, and J. Timur, "Prediksi Harga Rumah Menggunakan Metode Regressi Lasso," *Prosiding Seminar Implementasi Teknologi Informasi dan Komunikasi*, vol. 3, no. 2, 2024, doi: 10.31284/p.semtik.2024-2.6240.

[9] C. P. S. Heru Supriyono, "Pemilihan Rumah Tinggal Menggunakan Metode Weighted Product," *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, vol. 1, no. 1, pp. 23–28, 2015.

[10] D. Y. Descania, "Prediksi Pertumbuhan Penduduk Di Kecamatan Baregbeg Kabupaten Ciamis Menggunakan Metode Artificial Neural Network," *Jurnal Informatika dan Teknologi Pendidikan*, vol. 2, no. 2, pp. 37–43, Dec. 2022, doi: 10.25008/jitp.v2i2.28.

[11] Satrio Junaidi, R. Valicia Anggela, and D. Kariman, "Klasifikasi Metode Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa dengan Algoritma Naïve Bayes, Random Forest, Support Vector Machine (SVM) dan Artificial Neural Nerwork (ANN)," *Journal of Applied Computer Science and Technology*, vol. 5, no. 1, pp. 109–119, Jun. 2024, doi: 10.52158/jacost.v5i1.489.

[12] A. Rahman, R. Muniyandi, K. T. Islam, and Md. M. Rahman, "Ovarian Cancer Classification Accuracy Analysis Using 15-Neuron Artificial Neural Networks Model," Dec. 2019.

[13] M. L. Mu'tashim, T. Muhayat, S. A. Damayanti, H. N. Zaki, and R. Wirawan, "Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression,"

Informatik : Jurnal Ilmu Komputer, vol. 17, no. 3, p. 238, 2021, doi: 10.52958/iftk.v17i3.3635.

[14] W. Pangestu, A. Widodo, and B. Rahayudi, "Prediksi Jumlah Kendaraan Bermotor Di Indonesia Menggunakan Metode Average-Based Fuzzy Time Series Models memperoleh gelar Sarjana Komputer Disusun oleh :," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 9, p. 964, 2018.