

IMPLEMENTASI ALGORITMA REGRESI LINEAR UNTUK MEMPREDIKSI HARGA EMAS

Andini Fitriyah Salsabilah¹⁾, Muhammad Sabili Nurilhaq²⁾, Putra Dwi Wira G. Y.³⁾, Achmad Arbi Hanafi⁴⁾, M. Daffa Afillah.⁵⁾

E-mail : ¹⁾21081010217@student.upnjatim.ac.id , ²⁾20081010101@student.upnjatim.ac.id ,

³⁾20081010103@student.upnjatim.ac.id , ⁴⁾20081010104@student.upnjatim.ac.id ,

⁵⁾23084010031@student.upnjatim.ac.id

^{1,2,3,4)}Informatika, Fakultas Ilmu Komputer, UPN Veteran Jawa Timur

⁵⁾Bisnis Digital, Fakultas Ilmu Komputer, UPN Veteran Jawa Timur

(Naskah masuk: 25 November 2024, diterima untuk diterbitkan: 31 Desember 2024)

Abstrak

Penelitian ini bertujuan untuk memprediksi harga emas dengan menggunakan beberapa variabel bebas, yaitu kurs silver (SLV), indeks S&P 500 (SPX), kurs pada bursa saham "The United States Oil Fund" (USO), dan kurs nilai Euro (EUR) terhadap dollar Amerika Serikat (USD). Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari "Gold Price Data" dengan total sebanyak 2290 observasi dan 7 kolom. Metode yang digunakan adalah regresi, yang merupakan teknik untuk membangun model prediktif berdasarkan nilai input yang diberikan. Hasil prediksi dievaluasi berdasarkan nilai Root Mean Square Error (RMSE), di mana nilai yang lebih kecil menunjukkan akurasi yang lebih baik. Hasil penelitian menunjukkan bahwa model single variable memiliki akurasi sebesar 73%, sementara model multi variable memiliki akurasi sebesar 84%. Untuk meningkatkan akurasi prediksi, penelitian ini merekomendasikan penggunaan model prediktif yang lain serta perbaikan dalam pembagian dataset untuk memastikan distribusi yang lebih representatif. Penelitian ini tidak hanya berkontribusi dalam memprediksi harga emas, tetapi juga dalam pengembangan model prediksi yang lebih akurat dengan memanfaatkan variabel ekonomi yang relevan.

Kata Kunci : *prediksi harga emas, regresi, kurs silver, indeks S&P 500, RMSE.*

1. PENDAHULUAN

Investasi merupakan suatu bentuk penanaman modal yang berfungsi sebagai tabungan untuk masa kini dan masa depan. Salah satunya adalah investasi emas. Emas sebagai produk investasi telah menarik minat yang besar. Sifat emas sebagai logam mulia memberikan nilai tambah yang menarik bagi orang karena tahan korosi, lunak, mudah dibentuk, dan tidak hanya digunakan sebagai investasi tetapi juga sebagai barang koleksi [1]. Emas merupakan logam mulia yang banyak dicari karena nilainya secara umum stabil dan harga pergramnya terus meningkat dari tahun ke tahun[2]. Menurut [3] Nilai emas dapat berubah seiring perkembangan zaman, dan ketahanannya terhadap laju inflasi membuatnya menjadi pilihan umum untuk investasi. Investasi emas terbagi menjadi dua jenis, yaitu investasi digital dan investasi fisik. Investasi digital dibedakan menjadi dua jenis: tabungan dan perdagangan emas (*gold trading*). *Gold trading* dilakukan secara *online* tanpa menggunakan emas fisik [2].

Umumnya, individu yang terlibat dalam investasi emas seringkali membeli emas saat harganya dianggap sudah mencapai titik terendah, dengan harapan dapat menjualnya kembali saat harga emas berada pada level yang tinggi. Pendekatan ini dapat menimbulkan masalah karena memiliki tingkat kesalahan yang tinggi, dimana harga emas yang sedang mengalami penurunan dapat terus merosot ke tingkat yang lebih rendah. Hal ini menjadi berisiko karena jika harga emas terus turun, pelaku investasi emas bisa menghadapi kerugian besar saat harus menjual emasnya dengan harga yang jauh lebih rendah dibandingkan dengan harga saat emas tersebut dibeli. Untuk menyelesaikan permasalahan tersebut, maka dibutuhkan mekanisme dalam memprediksi harga emas. Salah satu algoritma yang dapat digunakan sebagai metode untuk melakukan prediksi adalah Regresi Linear [1]. Regresi linier adalah metode prediksi yang menggunakan suatu garis lurus untuk merepresentasikan keterkaitan antara dua variabel atau lebih [2]. Metode ini memiliki keunggulan nilai akurasi yang cukup tinggi dan lebih efisien dalam pengerjaannya[1].

Berdasarkan uraian di atas, pada penelitian ini menggunakan algoritma Regresi Linear yang bertujuan untuk membandingkan hasil prediksi harga emas di masa mendatang. Dengan menggunakan pendekatan ini, diharapkan para investor dapat memperoleh wawasan yang lebih akurat mengenai tren dan fluktuasi harga emas, sehingga dapat membuat keputusan investasi yang lebih cerdas. Dengan demikian, penerapan metode Regresi Linear diharapkan dapat menjadi sarana yang efektif bagi para investor untuk mengelola risiko dalam investasi emas mereka.

2. METODOLOGI

2.1 Pengumpulan Data

Proses pengumpulan data bertujuan agar mendapatkan informasi dari sumber yang valid, seperti: jurnal ilmiah, repositori, website, buku, dataset, dan sebagainya. Proses Pengumpulan data adalah sebuah aktivitas di mana informasi dikumpulkan melalui berbagai metode seperti observasi, kuesioner, wawancara, atau studi pustaka [4]. Pada penelitian ini, akan menggunakan data sekunder, yaitu data yang diperoleh secara tidak langsung dari objek atau subjek penelitian. Data penelitian ini didapatkan dari situs penyedia data Kaggle. Data yang digunakan yaitu data “*Gold Price Data*” dengan total data sebesar 2290 dengan kolom sebanyak 7.

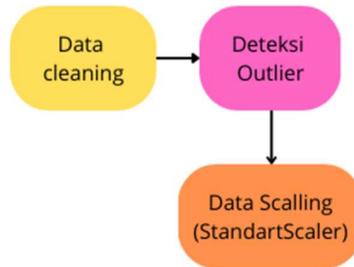
Data yang dibutuhkan pada penelitian ini berupa detail harga emas. Representasi ini memiliki berbagai macam bentuk misalnya berbentuk tabel dalam basis data, bentuk matriks, bentuk teks, bentuk *Comma Separated Value* (CSV) dan lain-lain [5]. Data yang digunakan berupa data berbentuk teks berformat .CSV. Ciri-ciri dataset yang baik adalah data yang lengkap, data yang *up to date*, konsistensi dalam representasi data, jumlah variabel yang jelas, tidak mengandung noise, menarik, serta mudah dimengerti [6].

Table 1. Atribut Data Nilai Kurs

Kolom	Jenis	Keterangan
DATE	Tanggal	Tanggal Kurs
SPX	Numerik	Indeks nilai pasar saham pada bursa saham AS
GLD	Numerik	Harga emas (<i>gold</i>)
USO	Numerik	Nilai tukar kurs pada bursa saham “ <i>The United State Oil Fund</i> ”
SLV	Numerik	Harga silver
EUR/USD	Numerik	Rasio nilai tukar kurs EUR terhadap kurs USD

2.2 Data Preprocessing

Data dapat dilakukan *training* apabila data telah dilakukan proses *preprocessing*[7]. Data *preprocessing* meliputi, data *cleaning* yaitu dibersihkan dari data kosong atau (*Null*), dan data duplikat, kemudian data *scaling* dengan penggunaan *StandardScaler* hal tersebut digunakan untuk mengkompres nilai data agar tidak terlalu besar dan mempermudah proses selanjutnya[8]. Namun pada proses penelitian sebelum *scaling* dilakukan deteksi *outlier* untuk memastikan data lebih bersih. Berikut adalah tahapan dalam data *preprocessing* :



Gambar 1. Alur Data Preprocessing

2.3 Model Regresi Linier

Regresi Linier merupakan metode statistika dalam membentuk model hubungan antara variabel terikat (*dependen*; respon; Y) dan variabel bebas (*independen*; prediktor; X [9]. Regresi adalah teknik dalam membangun model untuk keperluan prediksi nilai *input* yang diberikan. Apabila variabel bebas memiliki jumlah 1, model regresi yang dapat digunakan adalah Regresi linier sederhana. Sedangkan, model regresi linier berganda digunakan apabila jumlah variabel bebas lebih dari 1. Persamaan regresi linier sederhana secara matematik memiliki bentuk seperti persamaan (1)

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_3X_3 \quad (1)$$

Pada persamaan (1) variabel Y merupakan variabel terikat, yang memiliki ketergantungan hasil pada variabel X (variabel bebas). Sedangkan variabel a merupakan suatu konstanta dan variabel b merupakan koefisien dari variabel X . Pada persamaan (2) dan (3) merupakan persamaan yang digunakan dalam menentukan besarnya konstanta pada variabel a dan b .

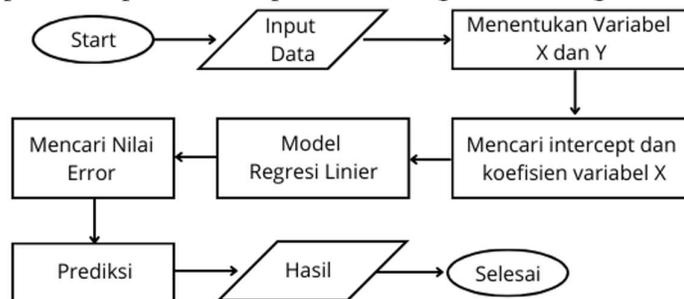
$$a = \frac{(\sum Y_i)(\sum X_i^2) - (\sum X_i)(\sum X_i Y_i)}{n\sum X_i^2 - (\sum X_i)^2} \quad (2)$$

$$b = \frac{n(\sum X_i Y_i) - (\sum X_i)(\sum Y_i)}{n\sum X_i^2 - (\sum X_i)^2} \quad (3)$$

catatan :

n = jumlah data

Berikut merupakan tahapan dalam implementasi dengan metode regresi linier[10]:



Gambar 2. Alur Implementasi Metode Regresi Linier

2.4 Root Mean Square Error (RMSE)

Perhitungan kesalahan memiliki guna mengevaluasi nilai hampiran, dimana semakin kecil nilai RMSE maka memiliki kecocokan dengan titik-titik data [11]. Hasil prediksi yang baik dapat ditentukan dengan akurasi nilai RMSE yang kecil [12]. Perhitungan RMSE dapat dilihat pada persamaan (4).

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (X_t - F_t)^2}{n}} \quad (4)$$

Catatan :

X_t = nilai aktual pada periode ke t

F_t = nilai peramalan pada periode ke t

$X_t - F_t$ = nilai kesalahan pada periode ke t

n = jumlah data.

3. HASIL DAN PEMBAHASAN

3.1 Pengolahan Data

Proses pengolahan data meliputi proses data *cleaning*, deteksi data *outlier*, dan *scaling*. Proses pengolahan *cleaning* data tidak ditemukan data duplikat dan nilai yang tidak ditemukan. Saat proses deteksi *outlier* ditemukan beberapa data *outlier* pada tiap kolom yang ditampilkan pada tabel 2 berikut :

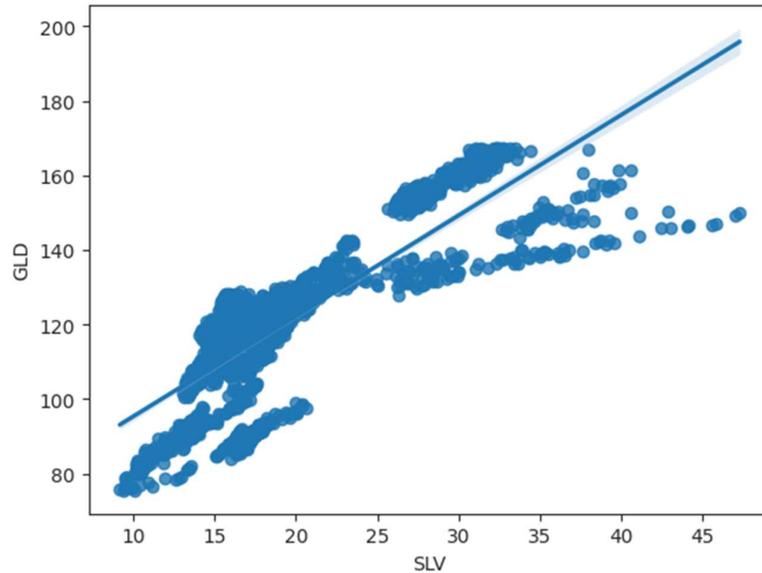
Table 2. Deteksi Data Outlier

Kolom	Jumlah Data
USO	134
SLV	117
EUR/USD	0
GLD	115

Dari paparan tabel diatas jumlah data sebelum dilakukan *preprocessing* memiliki jumlah sebesar 2290. Setelah dilakukan data *preprocessing* maka besar jumlah data yang dapat dilakukan *training* dan *testing* memiliki jumlah 1994.

3.2 Data Training

Proses *training* data dilakukan untuk mengetahui rentang percobaan prediksi nilai variabel Y terhadap variabel X. Training data yang dilakukan selama uji coba ditunjukkan pada gambar 3.

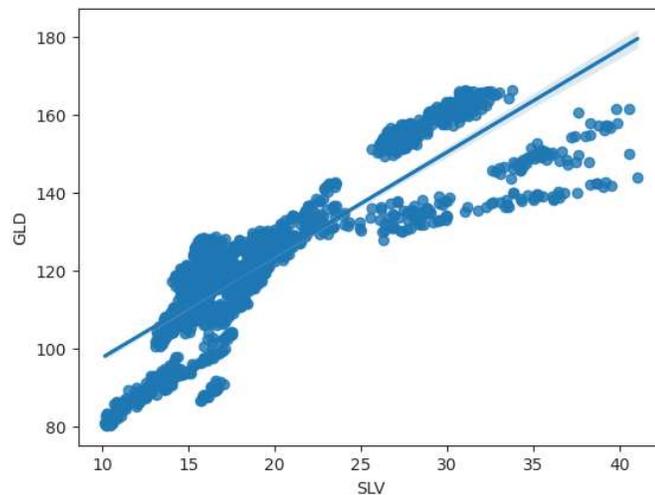


Gambar 3. Grafik Regresi Linier pada Data Training

Pada grafik diatas merupakan *training* data antar kurs silver sebagai sumbu X dan kurs emas sebagai sumbu Y. Grafik diatas menunjukkan bahwa data *Gold Price* memiliki kecocokan apabila dilakukan prediksi dengan metode regresi linier. Hal tersebut dikarenakan hubungan antar variabel kurs Gold dengan kurs Silver adalah linier.

3.3 Uji Coba Single Variable

Uji coba satu variabel meliputi variabel X yaitu kurs silver dan variabel Y yaitu kurs emas. Langkah pertama dalam melakukan uji coba dilakukan data *training* guna menentukan nilai *intercept* dan koefisien dari model regresi linier. Perhitungan nilai *intercept* dan koefisien dilakukan sesuai dengan persamaan (2) dan (3). Hasil perhitungan didapatkan nilai *intercept* sebesar 122.756731 dan nilai koefisien sebesar 16.457768. Pada uji coba dilakukan pembagian data yaitu data latih dan data test yang memiliki perbandingan 80 : 20. Nilai dari hasil tersebut akan dialihkan dalam bentuk grafik pada gambar 3.1.



Gambar 3.1. Grafik Regresi Linier pada Data Training Single Variable

Proses selanjutnya merupakan *testing* data. Pengujian dilakukan guna mendapatkan nilai akurasi dari model. Nilai akurasi yang dihasilkan menunjukkan nilai sebesar 73%.

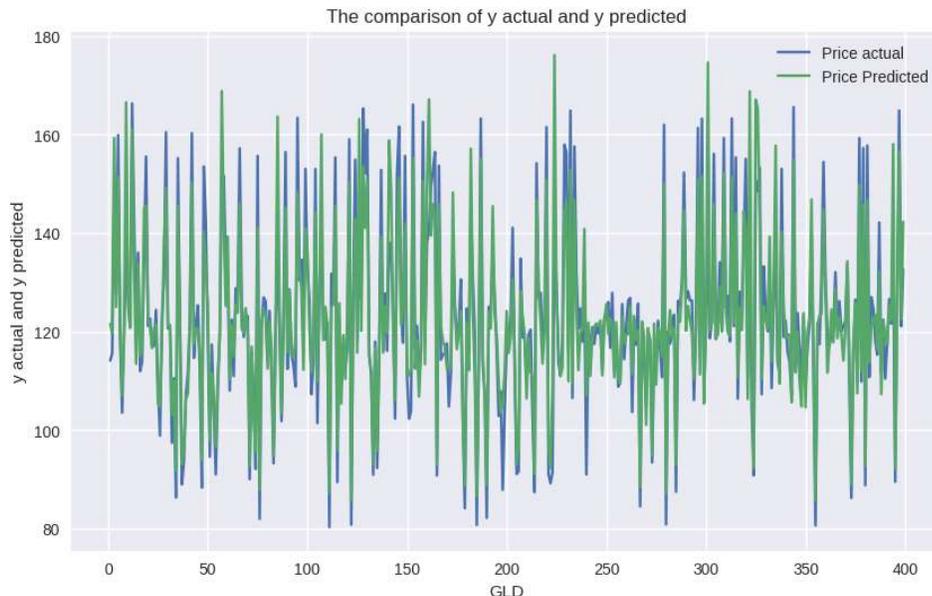
3.4 Uji Coba Multivariable

Selain implementasi regresi linear dengan satu variabel pada studi kasus ini, penelitian ini juga menerapkan regresi linear dengan multi variabel guna untuk mencari nilai akurasi yang tinggi. Pada pengujian multi variabel ini pembagian dataset menjadi data latih dan data uji sendiri yaitu 80:20, dan terdapat beberapa fitur yang tidak dilibatkan dalam pelatihan model ini. Fitur-fitur yang tidak diikutkan atau dengan kata lain tidak berpengaruh terhadap harga emas yaitu fitur DATE serta fitur Y (GLD) itu sendiri. Selanjutnya pada pengujian kali ini memperoleh nilai *coefisien* yang ditunjukkan pada Gambar 3.2 dan nilai *intercept* 122,814.

	Slope
EUR/USD	-2.051442
USO	0.059728
SPX	4.982692
SLV	18.775644

Gambar 3.2. Nilai Coefisien Multivariabel

Setelah membagi dataset dan telah mendapatkan nilai *coefisien* dan *intercept* maka selanjutnya model dilatih dengan data latih dengan menggunakan fungsi “fit()”. Selanjutnya yaitu *testing* dengan data uji yang telah ditentukan agar bisa memperoleh nilai akurasi dari model tersebut. Hasil dari *testing* model tersebut ditunjukkan pada Gambar 3.3 dan nilai akurasi yang diperoleh dalam pengujian regresi linear multivariabel ini yaitu bernilai 84%.



Gambar 3.3 Grafik Hasil Prediksi Multivariabel

3.5 Matriks Error

Testing dan *training* data dilakukan guna menghitung beberapa nilai galat seperti *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, nilai maksimum *error*, dan *R2 score*. Perhitungan *error* dilakukan dengan dilakukan pengukuran galat antara *output actual* dengan *output target*. Tahapan dalam perhitungan

galat dilakukan dengan *Sum Square Error* yang didapatkan dari penjumlahan nilai kuadrat error neuron 1 dan 2 pada tiap lapisan output data. Total penjumlahan *Sum Square Error* dibutuhkan guna menghitung nilai *Root Mean Squared Error (RMSE)* pada tiap iterasi [1]. Tabel 3 merupakan hasil perhitungan nilai galat pada proses *testing single variable*. Tabel 4 merupakan hasil perhitungan nilai galat pada proses *training single variable*. Tabel 5 merupakan hasil perhitungan nilai galat pada proses *testing multivariable*. Tabel 6 merupakan hasil perhitungan nilai galat pada proses *training multivariable*.

Tabel 3. Nilai Galat pada Proses Testing Single Variable

Keterangan	Hasil
<i>Mean Absolute Error (MAE)</i>	6.469994637715179
<i>Mean Squared Error (MSE)</i>	62.27199534068499
<i>Root Mean Squared Error (RMSE)</i>	7.89126069400099
Maksimum Error	27
R2 Score	0.82511

Tabel 4. Nilai Galat pada Proses Training Single Variable

Keterangan	Hasil
<i>Mean Absolute Error (MAE)</i>	6.581660316936940
<i>Mean Squared Error (MSE)</i>	66.43749527369637
<i>Root Mean Squared Error (RMSE)</i>	8.150919903526004
Maksimum Error	27
R2 Score	0.8179

Tabel 5. Nilai Galat pada Proses Testing Multi Variable

Keterangan	Hasil
<i>Mean Absolute Error (MAE)</i>	6.581660316936940
<i>Mean Squared Error (MSE)</i>	66.43749527369637
<i>Root Mean Squared Error (RMSE)</i>	8.150919903526004
Maksimum Error	27
R2 Score	0.8179

Tabel 6. Nilai Galat pada Proses Training Multi Variable

Keterangan	Hasil
<i>Mean Absolute Error (MAE)</i>	6.581660316936940
<i>Mean Squared Error (MSE)</i>	66.43749527369637
<i>Root Mean Squared Error (RMSE)</i>	8.150919903526004
Maksimum Error	27
R2 Score	0.8179

4. KESIMPULAN DAN SARAN.

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa terdapat tiga model berbeda dari algoritma regresi linear yang digunakan untuk prediksi harga emas. Model pertama menggunakan regresi linear dengan satu fitur yang dianggap sebagai faktor penentu, menghasilkan akurasi sebesar 73%. Model kedua, yang menambahkan variabel baru yakni X2 dan X3, meningkatkan akurasi menjadi 82%. Model ketiga, yang menggunakan multivariabel pada studi kasus kali ini, menunjukkan akurasi terbaik dibandingkan kedua model sebelumnya dengan nilai 84%. Dari kesimpulan tersebut, beberapa saran yaitu untuk meningkatkan akurasi lebih lanjut, penelitian selanjutnya dapat melakukan eksperimen dengan pembagian atau split dataset yang berbeda. Selain itu, dapat juga mempertimbangkan penggunaan algoritma regresi lainnya yang mungkin dapat meningkatkan akurasi prediksi.

5. DAFTAR RUJUKAN

- [1] Utomo, P. B., Utami, E., & Raharjo, S., 2019. Implementasi Metode K-Nearest Neighbor Dan Regresi Linear Dalam Prediksi Harga Emas. *Informasi Interaktif*, 4 (3), pp.155-159.
- [2] Aulia, A., Aprianti, B., Supriyanto, Y., & Rozikin, C., 2022. Prediksi Harga Emas dengan Menggunakan Algoritma Support Vector Regression (Svr) dan Linear Regression (LR). *Jurnal Ilmiah Wahana Pendidikan*, 8(5), pp.84-88.
- [3] Riyanto, V., Hamid, A., & Ridwansyah, R., 2019. Prediction of Student Graduation Time Using the Best Algorithm. *Indonesian Journal of Artificial Intelligence and Data Mining*, 2(1), pp.1-9.
- [4] A. N. Sihananto, A. Puspita Sari, H. Khariono, R. Akhmad Fernanda, and D. Cakra Mudra Wijaya, "Implementasi Metode K-Means Untuk Pengelompokan Kasus Covid-19 Tingkat Provinsi Di Indonesia," *J. Inform. dan Sist. Inf.*, vol. 3, no. 1, pp. 76–85, 2022, doi: 10.33005/jifosi.v3i1.472.
- [5] Rachmat, Antonius, and Yuan Lukito. "Sentipol: Dataset Sentimen Komentar Pada Kampanye Pemilu Presiden Indonesia 2014 Dari Facebook Page." *Konferensi Nasional Teknologi Informasi dan Komunikasi 2017 (2016)*: 218-228.
- [6] Hosseinkhah, Fatemeh, Hassan Ashktorab, and Ranjit Veen. "Challenges in data mining on medical databases." *Database technologies: Concepts, methodologies, tools, and applications*. IGI global, 2009. 1393-1404.
- [7] Çetin, V., & Yıldız, O. (2022). A comprehensive review on data preprocessing techniques in data analysis. *Pamukkale University Journal of Engineering Sciences*, 28(2), 299–312. <https://doi.org/10.5505/pajes.2021.62687>
- [8] S. I. Attaqwa, A. A. Hanafi, H. Hakim, A. A. Sofyan, and A. P. Sari, "Customer Clustering Menggunakan K-Means Agglomerative pada Pendapatan dan Pembelian daging," vol. 3, pp. 96–100, 2023.
- [9] Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. (2021). Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 7(1), 8–17. <https://doi.org/10.25077/teknosi.v7i1.2021.8-17>
- [10] N. Almumtazah, N. Azizah, Y. L. Putri, and D. C. R. Novitasari, "Prediksi Jumlah Mahasiswa Baru Menggunakan Metode Regresi Linier Sederhana," *J. Ilm. Mat. Dan Terap.*, vol. 18, no. 1, pp. 31–40, 2021, doi: 10.22487/2540766x.2021.v18.i1.15465.
- [11] Cahyani, H. C., Hidayah, E., Wiyono, R. U. A., Halik, G., & Widiarti, W. Y. (2021). Prediksi Laju Sedimentasi Pada Sungai Jatiroto. *Jurnal Rekayasa Sipil (JRS-Unand)*, 17(1), 64. <https://doi.org/10.25077/jrs.17.1.64-71.2021>

- [12] Indahyanti, U., Azizah, N. L., & Setiawan, H. (2022). Pendekatan ensemble learning untuk meningkatkan akurasi prediksi kinerja akademik mahasiswa. *Jurnal Sains Dan Informatika*, 8(2).