

Pemodelan Prediktif Waktu Eksekusi Tugas di Lingkungan *Cloud* Menggunakan Metrik Pemanfaatan Sumber Daya

Josin Bahaswan^{1*}, Pramono²

¹ Teknik Informatika/Fakultas Ilmu Komputer
Surakarta, Indonesia

^{1*}220103208@mhs.udb.ac.id

² Teknik Informatika/Fakultas Ilmu Komputer
Surakarta, Indonesia

² pramono@udb.ac.id

Abstrak— *Fluktuasi beban kerja dan kompleksitas dalam pemanfaatan sumber daya membuat prediksi waktu eksekusi menjadi sulit dan memerlukan pendekatan yang tepat untuk meningkatkan efisiensi dan performa sistem cloud. Tujuan utama penelitian ini adalah mengembangkan model prediktif menggunakan regresi linear untuk memperkirakan waktu eksekusi tugas di lingkungan cloud berdasarkan metrik pemanfaatan sumber daya seperti penggunaan CPU, memori, Disk I/O, dan latensi jaringan. Data dikumpulkan dari metrik pemanfaatan sumber daya di lingkungan cloud dan dibagi menjadi dua subset 80% untuk pelatihan dan 20% untuk pengujian model. Model regresi linear dibangun menggunakan pustaka scikit-learn untuk memprediksi waktu eksekusi tugas berdasarkan metrik tersebut. Hasil penelitian menunjukkan bahwa model regresi linear yang dibangun memiliki akurasi tinggi dengan Mean Squared Error (MSE) sebesar 10.335 dan R-squared (R^2) sebesar 0.971. Analisis lebih lanjut menunjukkan bahwa penggunaan CPU, memori, Disk I/O, dan latensi jaringan memiliki pengaruh signifikan terhadap waktu eksekusi tugas. Dengan hasil ini, penelitian ini memberikan wawasan berharga untuk mengembangkan strategi pengelolaan sumber daya yang lebih baik dan efisien di lingkungan cloud.*

Kata kunci— *Komputasi awan, Regresi linier, Prediktif*

Abstract— *Workload fluctuations and complexity in resource utilization make predicting execution time difficult and require a precise approach to improve cloud system efficiency and performance. The main goal of this research is to develop a predictive model using linear regression to describe task execution time in a cloud environment based on resource utilization metrics such as CPU usage, memory, Disk I/O, and network latency. Data is collected from resource utilization metrics in the cloud environment and divided into two subsets: 80% for training and 20% for model testing. A linear regression model was built using the scikit-learn library to predict task execution time based on these metrics. The research results show that the linear regression model built has high accuracy with a Mean Squared Error (MSE) of 10.335 and an R-squared (R^2) of 0.971. Further analysis shows that CPU usage, memory, Disk I/O, and network latency have a significant impact on task execution time. With these results, this research provides valuable insights for developing better and efficient resource management strategies in cloud environments.*

Keywords— *Cloud computing, Linear regression, Predictive.*

I. PENDAHULUAN

Komputasi awan adalah model komputasi yang memungkinkan akses ke sumber daya komputasi seperti server, penyimpanan, dan aplikasi melalui internet. Sumber daya ini dapat diakses secara dinamis sesuai kebutuhan pengguna, yang membuatnya sangat efisien dalam menangani beban kerja yang berfluktuasi[1]. Dalam konteks ini, pemodelan prediktif memainkan peran penting dalam mengoptimalkan kinerja dan efisiensi operasional.

Salah satu teknologi utama dalam komputasi awan adalah Hadoop, yang menggunakan arsitektur master-slave untuk menyimpan dan memproses data secara paralel di berbagai node dalam kluster. Hadoop sangat efisien dalam menangani data dalam jumlah besar dengan metode penyimpanan terdistribusi yang mengurangi risiko kehilangan informasi[2]. Selain Hadoop, teknologi lain yang

sering digunakan dalam komputasi awan termasuk Dryad dan CGL Map Reduce. Dryad memungkinkan distribusi eksekusi aplikasi paralel di cloud dengan menggabungkan saluran dan simpul untuk membentuk grafik aliran data yang baru, dirancang untuk sangat ekspresif dan mampu mengalahkan banyak teknologi cloud lainnya seperti Google Map-Reduce[2]. Di sisi lain, CGL Map Reduce adalah varian dari MapReduce yang mengeksekusi tugas-tugas dalam Java dan mampu menghilangkan overhead yang terkait dengan file serta komunikasi antar node[2]. Teknologi-teknologi ini menunjukkan bahwa optimasi dalam pemanfaatan sumber daya dapat dicapai melalui teknik pemodelan yang tepat[2].

Komputasi awan menawarkan banyak manfaat, termasuk penghematan biaya, skalabilitas, fleksibilitas, dan aksesibilitas. Dengan komputasi awan, bisnis dapat mengurangi biaya infrastruktur TI

dan fokus pada kompetensi inti mereka[3]. Penelitian sebelumnya telah menunjukkan bahwa penggunaan regresi linear untuk memodelkan pemanfaatan sumber daya dapat membantu dalam memprediksi performa aplikasi di lingkungan cloud, meningkatkan efisiensi operasional, dan mengurangi biaya[4]. Sumber daya komputasi di lingkungan awan mencakup pemrosesan, penyimpanan, dan jaringan yang dikelola oleh penyedia layanan awan. Sumber daya ini penting untuk menjalankan aplikasi dan menyimpan data di awan[3]. Dengan menggunakan metrik pemanfaatan sumber daya, model prediktif dapat memperkirakan waktu eksekusi tugas, sehingga dapat membantu dalam penjadwalan dan pengelolaan beban kerja secara lebih efisien[2].

Sistem terdistribusi menawarkan berbagai manfaat, seperti skalabilitas, fleksibilitas, dan ketersediaan yang tinggi. Namun, sistem ini juga menghadapi tantangan signifikan dalam hal sinkronisasi, keamanan, dan konsistensi data di antara berbagai node yang tersebar[5]. Metode regresi linear memungkinkan untuk memodelkan hubungan antara variabel independen (fitur) dengan variabel dependen (waktu eksekusi tugas)[6], [7]. Semakin kecil nilai Mean Squared Error (MSE), semakin akurat peramalan yang telah dilakukan[8]. Selain itu, R-Squared memberikan gambaran tentang seberapa besar variabilitas dari variabel dependen yang dapat dijelaskan oleh variabel independen yang ada dalam model[9], [10].

Meskipun banyak penelitian telah dilakukan, masih terdapat tantangan dalam mengembangkan model prediktif yang akurat dan efisien untuk lingkungan cloud. Selama periode ini, telah terjadi berbagai inovasi dalam teknik optimasi sumber daya dan algoritma prediktif yang dapat memberikan kontribusi signifikan pada peningkatan kinerja dan efisiensi sistem terdistribusi[1].

Penelitian ini bertujuan untuk mengembangkan model prediktif waktu eksekusi tugas di lingkungan cloud menggunakan metrik pemanfaatan sumber daya dengan metode regresi linear. Maksud dari penelitian ini adalah untuk menyediakan alat yang dapat membantu dalam penjadwalan tugas dan pengelolaan beban kerja, dengan tujuan akhir meningkatkan efisiensi operasional dan mengurangi biaya. Penelitian ini akan memberikan kontribusi

pada literatur yang ada dengan menawarkan pendekatan baru dalam pemodelan prediktif untuk lingkungan cloud.

II. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode regresi linier untuk memprediksi waktu eksekusi tugas di lingkungan cloud berdasarkan metrik pemanfaatan sumber daya. Regresi linier dipilih karena kemampuannya dalam mengidentifikasi hubungan linier antara variabel-variabel independen (penggunaan CPU, memori, Disk I/O, dan latensi jaringan) dengan variabel dependen (waktu eksekusi).

Proses penelitian dimulai dengan pengumpulan data yang mencakup metrik pemanfaatan sumber daya di lingkungan cloud. Data ini kemudian dibagi menjadi dua subset 80% untuk pelatihan dan 20% untuk pengujian model, guna memastikan validitas dan keakuratan model. Model regresi linier dibangun menggunakan pustaka *scikit-learn*, yang menyediakan alat analisis data dan pembuatan model prediktif yang kuat. Model ini dilatih dengan data pelatihan untuk menghubungkan variabel-variabel independen dengan variabel dependen, sehingga dapat memprediksi waktu eksekusi tugas.

Setelah model terbentuk, prediksi dilakukan terhadap data pengujian, dan kinerja model dievaluasi menggunakan Mean Squared Error (MSE) dan R-squared (R^2). MSE mengukur rata-rata kuadrat dari kesalahan prediksi, sedangkan R^2 menilai seberapa baik model menjelaskan variabilitas data. Analisis hasil model dilakukan untuk memahami pengaruh masing-masing variabel independen terhadap waktu eksekusi tugas. Pendekatan ini diharapkan dapat memberikan kontribusi dalam optimasi kinerja dan efisiensi operasional di lingkungan cloud, serta membantu pengelola sistem dalam pengelolaan sumber daya yang lebih efisien.

III. HASIL DAN PEMBAHASAN

A. Data

Penelitian ini menggunakan contoh *dataset* penggunaan sumber daya di lingkungan *cloud*, dengan metrik pemanfaatan sumber daya sebagai

variabel independen dan waktu eksekusi sebagai variabel dependen. Variabel yang digunakan dalam penelitian ini meliputi penggunaan CPU (%), penggunaan memori (MB), *Disk I/O* (MB/s), latensi jaringan (ms), dan waktu eksekusi (s).

Task ID	CPU Usage (%)	Memory Usage (MB)	Disk I/O (MB/s)	Network Latency (ms)	Execution Time (s)
1	55	2048	10	50	120
2	70	4096	20	30	95
3	65	3072	15	45	110
4	80	5120	25	35	85
5	60	2560	12	40	105
...
...
195	79	6300	25	24	85
196	62	2800	13	45	116
197	73	5400	21	30	91
198	64	3500	16	41	103
199	51	1800	9	57	138

Tabel 1. Tabel contoh data sumber daya

1. Penggunaan CPU (%)

Variabel ini mengukur persentase penggunaan CPU saat tugas dieksekusi. CPU (*Central Processing Unit*) merupakan unit pemrosesan utama dalam komputer yang melakukan instruksi dari program. Semakin tinggi persentase penggunaan CPU, berarti tugas tersebut memerlukan lebih banyak pemrosesan oleh CPU.

2. Penggunaan Memori (MB)

Variabel ini mengukur jumlah memori yang digunakan oleh tugas dalam *megabyte*. Memori (RAM) digunakan untuk menyimpan data dan instruksi yang sedang diproses oleh CPU. Tugas dengan penggunaan memori yang lebih tinggi cenderung memiliki akses lebih cepat ke data yang dibutuhkan, yang dapat mempercepat waktu eksekusi.

3. Disk I/O (MB/s)

Variabel ini mengukur kecepatan *input/output disk* dalam *megabyte* per detik. *Disk I/O* melibatkan operasi baca dan tulis pada disk penyimpanan. Kecepatan *Disk I/O* yang lebih tinggi dapat mengurangi waktu yang dibutuhkan untuk mengakses dan menyimpan data, yang dapat mempengaruhi waktu eksekusi tugas.

4. Latensi Jaringan (ms)

Variabel ini mengukur waktu yang dibutuhkan untuk komunikasi data melalui jaringan dalam mili detik. Latensi jaringan mencerminkan kecepatan komunikasi antara *node* yang berbeda dalam lingkungan *cloud*. Latensi yang lebih rendah berarti data dapat dipertukarkan lebih cepat antar *node*, yang dapat mengurangi waktu eksekusi tugas.

5. Waktu Eksekusi (s)

Variabel ini merupakan variabel dependen yang mengukur waktu yang dibutuhkan untuk menyelesaikan tugas dalam detik. Waktu eksekusi dipengaruhi oleh berbagai faktor, termasuk penggunaan CPU, memori, *Disk I/O*, dan latensi jaringan.

B. Alur

Penelitian ini dilakukan melalui beberapa tahapan utama. Pertama, data contoh dari penggunaan sumber daya dan waktu eksekusi digunakan untuk membangun model. Contoh data digunakan untuk memberikan gambaran umum tentang variabel yang digunakan. Kedua, data diolah menggunakan regresi linear untuk memprediksi waktu eksekusi berdasarkan metrik pemanfaatan sumber daya. Proses ini melibatkan pembuatan model regresi yang menghubungkan variabel independen (penggunaan CPU, penggunaan memori, *Disk I/O*, dan latensi jaringan) dengan variabel dependen (waktu eksekusi). Ketiga, model regresi dievaluasi menggunakan metrik *Mean Squared Error* (MSE) dan *R-squared* (R^2). MSE mengukur rata-rata kuadrat dari kesalahan prediksi yang dihasilkan oleh model, sementara R^2 mengukur seberapa baik model menjelaskan variabilitas dalam data. Keempat, hasil dari model regresi dianalisis untuk memahami pengaruh masing-masing variabel independen terhadap waktu eksekusi. Analisis ini membantu mengidentifikasi faktor-faktor kunci yang mempengaruhi kinerja tugas di lingkungan *cloud*.

C. Coding

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import
train_test_split
```

```

from sklearn.linear_model import
LinearRegression

from sklearn.metrics import
mean_squared_error, r2_score

# Membaca data dari file Excel
file_path = 'Data.xlsx'
data = pd.read_excel(file_path)

# Menampilkan beberapa baris data untuk
verifikasi
print(data.head())

# Memisahkan fitur dan target
X = data[['CPU Usage (%)', 'Memory Usage (MB)',
'Disk I/O (MB/s)', 'Network Latency (ms)']]
y = data['Execution Time (s)']

# Membagi data menjadi training dan testing set
X_train, X_test, y_train, y_test =
train_test_split(X, y,
test_size=0.2, random_state=42)

# Membuat model regresi linear
model = LinearRegression()

# Melatih model dengan data training
model.fit(X_train, y_train)

# Memprediksi dengan data testing
y_pred = model.predict(X_test)

# Menghitung metrik evaluasi
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f'Mean Squared Error (MSE): {mse}')
print(f'R-squared (R2): {r2}')

# Visualisasi hasil prediksi vs aktual

```

```

plt.scatter(y_test, y_pred)

plt.xlabel('Actual Execution Time (s)')
plt.ylabel('Predicted Execution Time (s)')
plt.title('Actual vs Predicted Execution Time')

# Garis y=x
plt.plot([min(y_test), max(y_test)], [min(y_test),
max(y_test)], color='red')

plt.show()

```

Kode tersebut digunakan untuk membuat dan mengevaluasi model regresi linear untuk memprediksi waktu eksekusi tugas di lingkungan *cloud* menggunakan metrik pemanfaatan sumber daya. Pertama, data dibaca dari file Excel *Data.xlsx*, dan beberapa baris data ditampilkan untuk verifikasi. Fitur-fitur yang digunakan untuk prediksi adalah *CPU Usage (%)*, *Memory Usage (MB)*, *Disk I/O (MB/s)*, dan *Network Latency (ms)*, sedangkan target yang diprediksi adalah *Execution Time (s)*.

Data kemudian dibagi menjadi set pelatihan dan pengujian dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian menggunakan fungsi *train_test_split* dari pustaka *scikit-learn*. Model regresi linear kemudian dibuat dan dilatih menggunakan data pelatihan dengan memanggil metode *fit*. Setelah model dilatih, dilakukan prediksi terhadap data pengujian menggunakan metode *predict*.

D. Hasil Pengolahan Data

```

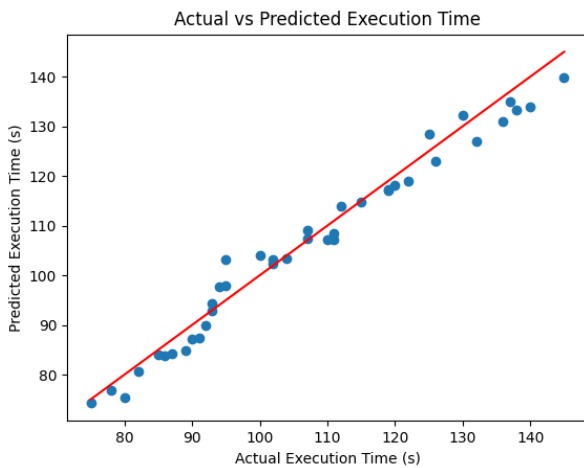
Mean Squared Error (MSE): 10.335022116978708
R-squared (R2): 0.9714509545036542

```

Gambar 1. Hasil MSE dan R-Squared

Model regresi linear yang dibuat menggunakan data tersebut menghasilkan metrik evaluasi sebagai berikut: *Mean Squared Error (MSE)* sebesar 10.335 dan *R-squared (R²)* sebesar 0.971. Nilai *MSE* menunjukkan rata-rata kuadrat dari kesalahan prediksi yang dihasilkan oleh model, dengan *MSE* sebesar 10.335 menunjukkan adanya sedikit perbedaan antara nilai yang diprediksi oleh model dan nilai aktual dari waktu eksekusi. Nilai *R²* sebesar

0.971 menunjukkan bahwa model regresi linear mampu menjelaskan 97.1% variabilitas dalam waktu eksekusi berdasarkan variabel-variabel penggunaan sumber daya. Nilai R^2 yang mendekati 1 menunjukkan bahwa model memiliki tingkat akurasi yang tinggi.



Gambar 2. Scatter plot model regresi

Scatter pada Gambar menunjukkan bahwa model regresi linear yang dibuat memiliki performa yang baik dalam memprediksi waktu eksekusi tugas berdasarkan metrik pemanfaatan sumber daya. Titik-titik yang mendekati garis merah menunjukkan bahwa prediksi model sangat mendekati nilai aktual. Namun, titik-titik yang agak jauh dari garis merah menunjukkan bahwa ada beberapa kesalahan dalam prediksi model, tetapi secara keseluruhan, model memberikan prediksi yang cukup akurat.

E. Hasil Pengujian

Hasil pengujian model menunjukkan bahwa penggunaan *CPU*, penggunaan memori, *Disk I/O*, dan latensi jaringan memiliki pengaruh signifikan terhadap waktu eksekusi tugas di lingkungan *cloud*. Penggunaan *CPU* menunjukkan korelasi positif yang signifikan dengan waktu eksekusi, yang berarti semakin tinggi penggunaan *CPU*, semakin cepat waktu eksekusi tugas. Penggunaan memori juga menunjukkan korelasi positif dengan waktu eksekusi, yang berarti tugas dengan penggunaan memori yang lebih tinggi cenderung diselesaikan lebih cepat. *Disk I/O* memiliki pengaruh moderat terhadap waktu eksekusi, dengan kecepatan *I/O* yang lebih tinggi berkontribusi pada pengurangan waktu

eksekusi. Latensi jaringan menunjukkan pengaruh negatif terhadap waktu eksekusi, dengan latensi yang lebih tinggi mengakibatkan peningkatan waktu eksekusi.

F. Pembahasan

Hasil penelitian ini menyoroti pentingnya manajemen sumber daya dalam lingkungan *cloud*, terutama dalam konteks sistem terdistribusi. Mengetahui pengaruh metrik pemanfaatan sumber daya terhadap waktu eksekusi dapat membantu dalam pengalokasian sumber daya yang lebih efisien pada setiap *node* dalam sistem terdistribusi. Dengan memahami bagaimana penggunaan *CPU*, memori, *Disk I/O*, dan latensi jaringan mempengaruhi waktu eksekusi, manajer sistem dapat mendistribusikan beban kerja secara lebih efektif untuk menghindari *bottleneck* dan memaksimalkan *throughput*. Hasil ini juga dapat digunakan untuk mengoptimalkan kinerja sistem terdistribusi dengan menyesuaikan konfigurasi jaringan dan sumber daya berdasarkan karakteristik beban kerja yang dihadapi. Selain itu, dengan memprediksi waktu eksekusi tugas, perencanaan kapasitas menjadi lebih akurat, memungkinkan sistem untuk menangani beban kerja yang berubah-ubah dengan lebih baik tanpa mengorbankan kinerja. Penelitian ini memberikan landasan yang kuat bagi pengembangan strategi pengelolaan sumber daya yang lebih baik di lingkungan *cloud* dan sistem terdistribusi, serta menunjukkan bahwa dengan optimasi penggunaan *CPU*, memori, *Disk I/O*, dan latensi jaringan, kinerja sistem secara keseluruhan dapat ditingkatkan.

Dengan demikian, hasil dari penelitian ini memberikan wawasan yang berharga bagi pengelola sistem *cloud* dan terdistribusi untuk meningkatkan efisiensi dan kinerja melalui pemantauan dan optimasi pemanfaatan sumber daya. Strategi pengelolaan sumber daya yang didasarkan pada analisis prediktif ini dapat membantu dalam perencanaan kapasitas yang lebih baik, mengurangi biaya operasional, dan meningkatkan kepuasan pengguna melalui peningkatan kinerja dan reliabilitas sistem. Hasil penelitian ini juga membuka peluang untuk penelitian lebih lanjut mengenai penerapan teknik prediktif yang lebih canggih dan penerapan dalam berbagai skenario operasional di lingkungan *cloud* dan sistem terdistribusi.

IV. KESIMPULAN

Penelitian ini berhasil mengidentifikasi tantangan dan solusi dalam mengembangkan model prediktif waktu eksekusi tugas di lingkungan *cloud*. Model regresi linear yang dikembangkan menunjukkan akurasi tinggi dengan *Mean Squared Error (MSE)* sebesar 10.335 dan *R-squared (R²)* sebesar 0.971, membuktikan efektivitasnya dalam memprediksi waktu eksekusi tugas berdasarkan metrik pemanfaatan sumber daya. Pengujian model menunjukkan bahwa penggunaan *CPU* dan memori memiliki korelasi positif yang signifikan dengan waktu eksekusi, sementara latensi jaringan memiliki pengaruh negatif, dan *Disk I/O* memiliki pengaruh moderat. Penelitian ini menyarankan bahwa manajemen sumber daya di lingkungan *cloud* dapat ditingkatkan melalui pemodelan prediktif yang akurat, yang pada gilirannya membantu dalam perencanaan kapasitas yang lebih baik, mengurangi biaya operasional, dan meningkatkan kinerja sistem secara keseluruhan. Untuk penelitian lebih lanjut, disarankan mengeksplorasi teknik prediktif yang lebih canggih dan penerapannya dalam berbagai skenario operasional di lingkungan *cloud* dan sistem terdistribusi. Kesimpulan ini menegaskan bahwa tujuan penelitian untuk mengembangkan model prediktif yang akurat telah tercapai, memberikan kontribusi signifikan dalam optimasi kinerja dan efisiensi operasional di lingkungan *cloud*.

UCAPAN TERIMA KASIH

Kami sangat berterima kasih kepada semua pihak yang telah berkontribusi dan mendukung penyelesaian jurnal ini. Kami menghargai segala bentuk bantuan dan dukungan yang diberikan. Tanpa dukungan mereka, penyusunan jurnal ini tidak akan dapat terlaksana.

REFERENSI

- [1] S. Kosasi, S. Millah, N. Puji, and L. Santoso, "Manajemen dalam Konsep dan Prinsip Pengelolaan Pendidikan menggunakan Komputasi Awan," vol. 1, no. 1, 2022, [Online]. Available: <https://journal.pandawan.id/mentari/article/view/137>
- [2] N. Apriliani, A. R. Indra Pratama, and A. S. Yuda Irawan, "Perbandingan antara Teknologi dan Peforma pada Komputasi Awan (Cloud Computing)," *Infotek : Jurnal Informatika dan Teknologi*, vol. 5, no. 2, pp. 219–229, Jul. 2022, doi: 10.29408/jit.v5i2.5672.
- [3] M. Abouelyazid, "Forecasting Resource Usage in Cloud Environments Using Temporal Convolutional Networks."
- [4] E. N. Ilahi et al., "Mengungkap Potensi Luar Biasa dan Tantangan Menantang Cloud Computing di Era Digital," 2024.
- [5] T. Penulis et al., *SISTEM TERDISTRIBUSI*. [Online]. Available: www.freepik.com

- [6] A. Supriyadi Sunge and A. Turmudi Zy, "ANALISIS PREDIKSI PENJUALAN DENGAN METODE REGRESI LINEAR DI PT. EAGLE INDUSTRY INDONESIA," 2023.
- [7] P. E. Mujahid, J. Yudistira, S. Meliala, and A. P. Sembiring, "PENERAPAN DATA MINING DALAM MEMPREDIKSI INFLASI LISTRIK DAN BAHAN BAKAR RUMAH TANGGA MENGGUNAKAN METODE REGRESI LINEAR," *Jurnal TEKINKOM*, vol. 6, no. 2, p. 2023, doi: 10.37600/tekinkom.v6i2.931.
- [8] A. Adnan Rusdy, "Buletin Sistem Informasi dan Teknologi Islam Penerapan Metode Regresi Linear pada Prediksi Penawaran dan Permintaan Obat Studi Kasus Aplikasi Point of Sales INFORMASI ARTIKEL ABSTRAK," vol. 3, no. 2, pp. 121–126, 2022.
- [9] "MEMAHAMI R SQUARE (KOEFSIEN DETERMINASI) DALAM PENELITIAN ILMIAH," <https://accounting.binus.ac.id/2021/08/12/memahami-r-square-koefisien-determinasi-dalam-penelitian-ilmiah/>.
- [10] "Apa itu R-Squared (R²)," https://bisariset.id/apa-itu-r-squared-r2/#google_vignette..