

Implementasi Algoritma Regresi Linier untuk Mengukur Tingkat Pengeluaran Co2 Pada Kendaraan Bermotor

Kevin Asgaryansyah

Jurusan Teknik Elektro, Universitas Mataram

E-mail: kevinfaizkf@gmail.com

Paniran

Jurusan Teknik Elektro, Universitas Mataram

Email: paniranmt@yahoo.com

***Abstract.** This study employs a linear regression algorithm to measure CO₂ emissions from motor vehicles based on the type of fuel used—gasoline, premium, diesel, and ethanol—as well as the distance traveled per kilometer. With growing concerns about the environmental impact of vehicle emissions, this research aims to develop a predictive model to estimate CO₂ emissions based on these variables. Despite the limited data used, the results indicate that the linear regression model has high accuracy in predicting CO₂ emissions. This model can serve as an effective tool to mitigate the environmental impact of motor vehicles and assist policymakers in designing better emission mitigation strategies*

Keywords: Linear Regression, CO₂ Emission, Fuel type, Prediction

Abstrak. Penelitian ini menggunakan algoritma regresi linier untuk mengukur tingkat pengeluaran CO₂ pada kendaraan bermotor berdasarkan jenis bahan bakar yang digunakan, yaitu bensin, premium, diesel, dan etanol, serta jarak tempuh kendaraan per kilometer. Dengan meningkatnya kekhawatiran tentang dampak emisi kendaraan terhadap lingkungan, penelitian ini bertujuan membangun model prediktif untuk memperkirakan emisi CO₂ berdasarkan variabel-variabel tersebut. Meskipun data yang digunakan terbatas, hasil penelitian menunjukkan bahwa model regresi linier memiliki akurasi tinggi dalam memprediksi emisi CO₂. Model ini dapat digunakan sebagai alat yang efektif dalam upaya mengurangi dampak lingkungan dari kendaraan bermotor dan membantu pembuat kebijakan dalam merancang strategi mitigasi emisi yang lebih baik.

Kata Kunci: Regresi linier, Emisi CO₂, Jenis bahan bakar, Prediksi.

PENDAHULUAN

Perubahan iklim global telah menjadi isu kritis yang mempengaruhi ekosistem, kesehatan manusia, dan ekonomi di seluruh dunia. Salah satu kontributor utama terhadap perubahan iklim adalah emisi gas rumah kaca, terutama karbon dioksida (CO₂), yang dihasilkan oleh berbagai aktivitas manusia, termasuk transportasi (Jainal et.al., 2022). Kendaraan bermotor merupakan sumber utama emisi CO₂, terutama di perkotaan yang padat lalu lintas. Oleh karena itu, pengukuran yang akurat dan efektif terhadap tingkat pengeluaran CO₂ dari kendaraan bermotor sangat penting untuk mengembangkan strategi mitigasi yang tepat (Dita et al., 2019).

Regresi linier adalah salah satu algoritma machine learning yang paling sederhana dan efisien untuk memprediksi nilai kontinu berdasarkan data historis (Harsiti et.al, 2022). Berbagai jenis bahan bakar memiliki karakteristik emisi yang berbeda. Bensin, premium, diesel, dan etanol adalah empat jenis bahan bakar yang umum digunakan pada kendaraan bermotor, masing-masing dengan profil emisi CO₂ yang unik (Sari et.al, 2021). Bensin adalah

bahan bakar yang paling umum digunakan dan menghasilkan emisi CO₂ yang signifikan karena pembakarannya yang tidak sempurna. Premium, sebagai varian bensin dengan oktan lebih tinggi, dapat sedikit mengurangi emisi tetapi tetap berkontribusi signifikan terhadap CO₂. Diesel, meskipun lebih efisien dalam konsumsi bahan bakar, menghasilkan emisi CO₂ yang tinggi dan polutan lain seperti NO_x. Etanol, sebagai bahan bakar alternatif yang lebih ramah lingkungan, menghasilkan emisi CO₂ yang lebih rendah dibandingkan dengan bensin dan diesel (A'yun et.al, 2023).

Penelitian ini bertujuan untuk mengimplementasikan algoritma regresi linier dalam mengukur tingkat pengeluaran CO₂ pada kendaraan bermotor. Dengan menggunakan data yang relevan dan teknik analisis statistik, penelitian ini akan mengembangkan model prediktif yang dapat digunakan oleh pemerintah, pembuat kebijakan, dan industri otomotif untuk mengurangi dampak lingkungan dari emisi kendaraan bermotor.

Melalui studi ini, diharapkan dapat ditemukan model yang tidak hanya akurat tetapi juga mudah diimplementasikan di berbagai konteks, membantu dalam upaya global untuk mengurangi emisi gas rumah kaca dan memerangi perubahan iklim. Penelitian ini juga akan menyoroti pentingnya pemahaman mendalam tentang faktor-faktor yang mempengaruhi emisi kendaraan, memberikan wawasan yang berharga untuk pengembangan kebijakan lingkungan yang lebih efektif.

LANDASAN TEORI

1. Regresi Linier

Regresi linier adalah salah satu teknik pembelajaran yang umum digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen (fitur) dan satu variabel dependen (target) yang bersifat kontinu. Tujuannya adalah untuk menghasilkan model prediktif yang dapat memprediksi nilai target berdasarkan nilai-nilai fitur yang diberikan (Theofani et al., 2019). Secara matematis, regresi linier mencoba untuk mempelajari koefisien atau bobot yang menghubungkan setiap fitur dengan target, sehingga memungkinkan model untuk membuat prediksi berdasarkan kombinasi linier dari fitur-fitur tersebut. Model regresi linier sering kali digunakan dalam kasus-kasus di mana hubungan antara fitur dan target dapat dijelaskan secara linier. Model regresi linier sederhana melibatkan satu variabel independen dan satu variabel dependen. Persamaan regresi linier yang digunakan adalah:

$$y = Ax + B \dots \dots \dots (1)$$

Keterangan:

y = variabel dependent (terikat)

x = variabel independent (bebas)

A = Kemiringan (slope)

B = konstanta (intercept)

Persamaan (1) merupakan persamaan regresi linier sederhana. Dalam konteks ini:

- y adalah variabel dependen atau terikat yang ingin diprediksi.
- x adalah variabel independen atau bebas yang digunakan untuk memprediksi y.
- A adalah kemiringan garis regresi, yang menunjukkan seberapa besar perubahan dalam y untuk setiap satu unit perubahan dalam x.
- B adalah konstanta atau intercept, yang merupakan titik di mana garis regresi memotong sumbu y saat x sama dengan nol.

Persamaan (1) menggambarkan hubungan linier antara x dan y, di mana A menentukan kemiringan dan B menentukan titik awal dari garis regresi. Dengan menggunakan teknik regresi linier, nilai-nilai dari A dan B dapat ditentukan melalui proses pembelajaran dari data yang diberikan. Setelah diperoleh, model regresi linier ini dapat digunakan untuk memprediksi nilai y untuk nilai-nilai x yang baru.

2. Explanatory Data Analysis (EDA)

Exploratory Data Analysis (EDA) adalah pendekatan analitis dalam statistik yang digunakan untuk menganalisis data dengan tujuan untuk meringkas karakteristik utamanya menggunakan metode visual dan statistic (Noviyanti et.al, 2022). EDA merupakan langkah awal yang esensial dalam proses analisis data dan pemodelan machine learning, karena membantu peneliti untuk memahami struktur data, menemukan pola, mendeteksi anomali, dan memeriksa asumsi yang mendasari data. Melalui EDA, peneliti dapat menggunakan statistik deskriptif untuk menghitung ukuran seperti mean, median, mode, variansi, dan standar deviasi untuk meringkas data, serta visualisasi data seperti histogram, box plot, scatter plot, dan pair plot untuk menggambarkan data secara visual (Wahyuni et.al, 2019). Selain itu, analisis korelasi juga digunakan untuk mengukur dan memvisualisasikan hubungan antar variabel.

3. Google Colaboratory

Google Colaboratory, atau Google Colab, adalah layanan berbasis cloud yang disediakan oleh Google untuk menulis dan menjalankan kode Python langsung di browser.

Colab memungkinkan pengguna untuk mengakses lingkungan komputasi yang kuat tanpa memerlukan perangkat keras khusus di komputer mereka. Salah satu fitur utamanya adalah akses gratis ke GPU dan TPU, yang sangat bermanfaat untuk tugas-tugas pemelajaran mesin dan analisis data yang membutuhkan komputasi tinggi (Soen et.al, 2022). Colab juga terintegrasi dengan Google Drive, memudahkan pengguna untuk menyimpan, mengakses, dan berbagi file. Selain itu, Colab menggunakan format notebook Jupyter, yang memungkinkan penggabungan kode, teks yang diformat dengan Markdown, dan visualisasi dalam satu dokumen interaktif. Fitur kolaborasi real-time ala Google Docs memungkinkan beberapa pengguna untuk bekerja bersama pada notebook yang sama, meningkatkan produktivitas dan kemudahan berbagi ide. Dengan berbagai pustaka Python populer yang sudah terinstal sebelumnya, seperti TensorFlow, Keras, dan PyTorch, Colab menjadi alat yang sangat efektif dan mudah diakses untuk pengembangan dan eksperimen dalam bidang data science dan machine learning (Guntara et.al, 2023).

METODOLOGI PENELITIAN

Sebelum dilakukannya penginputan data pada model, perlu dilakukan pendekatan analitis statistic atau biasa yang disebut dengan EDA untuk menganalisis dataset dan meringkas karakteristik utama dari dataset, sehingga dataset siap untuk digunakan, dataset yang digunakan adalah Fuel Consumption Ratings 2023 diambil dari Kaggle.com. Berikut angka yang harus dilakukan:

1. Data Understanding

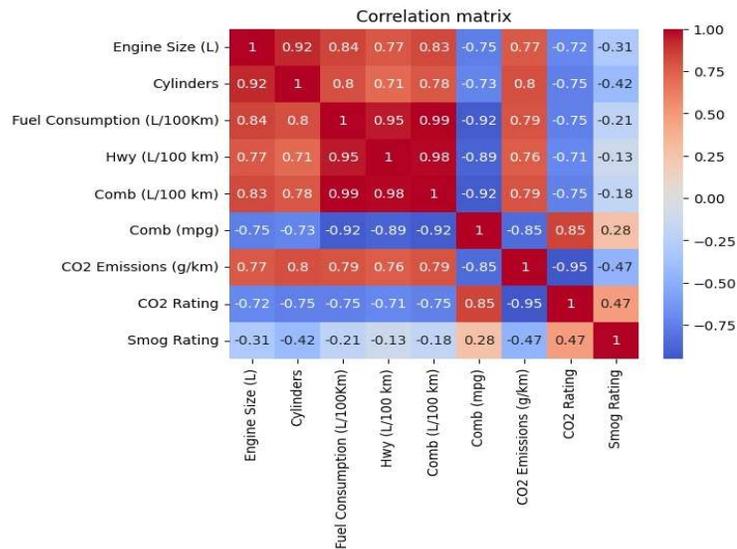
Tahap pertama adalah data understanding, untuk memahami isi secara keseluruhan dari dataset dengan melihat berapa banyak jumlah baris dan kolom dan melihat rata – rata dari persebaran data yang ada.

Tabel 1. Isi Dataset secara keseluruhan

	Engine size	Cylinder	Fuel consumption	Highway	Comb	CO2 Emission	Smog Rating	Fuel Type
1	1.6	6.0	6.4	7.0	6.6	157	7.0	Premium
2	1.5	6.0	9.1	7.6	8.4	197	6.0	Etanol
3	5.7	6.0	14.7	9.4	12.3	289	3.0	Diesel
...
1255	5.3	6.0	19.7	12.4	15.2	253	3.0	Premium
Mean	3.347	5.97	13.15	9.97	11.73	262	4.80	

2. Data Preparation

Tahap berikutnya adalah data preparation, dimana data yang telah diamati, akan di seleksi berdasarkan tingkat korelasi antar variable. Nilai 1 berkolerasi, dan 0 sebaliknya.



Berdasarkan gambar 1, ada beberapa variabel yang memiliki tingkat korelasi yang tinggi dengan CO2 Emission, diantaranya cylinder, comb (kombinasi penggunaan bahan bakar dijalanan raya), dan engine size. Karena penelitian ini lebih berfokus kepada jumlah emisi CO2 yang dikeluarkan, maka jarak kendaraan bermotor sangat diperlukan untuk mengukur berapa banyak tingkat emisi yang dikeluarkan, sehingga variabel comb digunakan sebagai inputan model yang digunakan, dengan output berupa CO2 Emission.

3. Feature Selection

Setelah dilakukannya data preparation, didapatkanlah dataset yang telah siap untuk digunakan sebagai data training dan data test untuk model yang akan digunakan, fitur yang digunakan sebagai input adalah comb, merupakan kependekan dari combination, yang merupakan gabungan penggunaan kendaraan bermotor di jalan raya dan pedesaan. Untuk outputnya menggunakan fitur CO2 emission dan fuel consumption, dimana terdapat 4 tipe bahan bakar yang digunakan, diantaranya etanol, premium, bensin, dan diesel yang masing masing diberi inputan yang sama. Berikut adalah tabel data yang telah di preparasi sebelumnya:

Tabel 2. Dataset yang telah di preparasi

No	Comb(Km)	CO2 Emission	Fuel Type
1	6.6	157	Premium
2	8.4	197	Etanol
3	12.3	259	Diesel
...
1254	17.7	293	Regular Gasoline
1255	15.2	318	Premium

4. Modelling

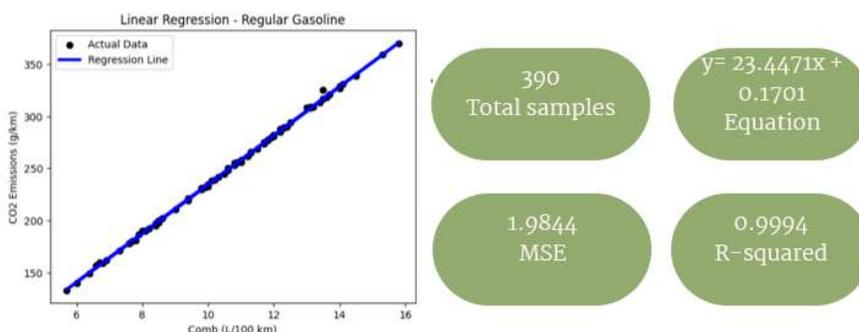
Dataset kemudian dimasukkan ke dalam model dengan membagi rasio 80% data training dan 20% data test. Kemudian data dilatih dengan menggunakan algoritma regresi linier dengan menggunakan persamaan sebagai berikut:

$$A = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \dots\dots\dots(2)$$

$$B = \frac{(\sum y)(\sum x)^2 - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \dots\dots\dots(3)$$

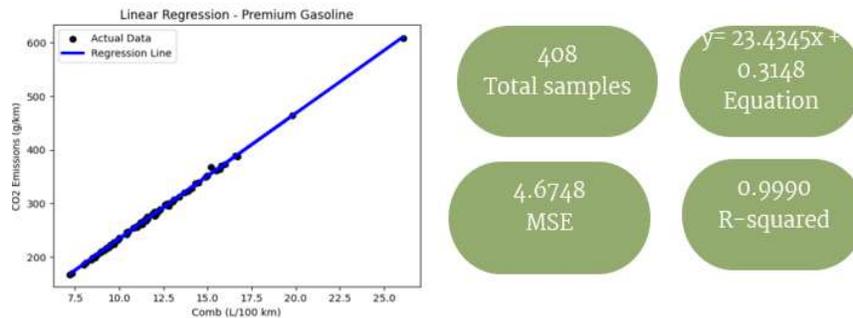
HASIL DAN PEMBAHASAN

Setelah dilakukannya pelatihan model, akan didapatkan hasil dari masing – masing tipe bahan bakar yang digunakan sebagai berikut:



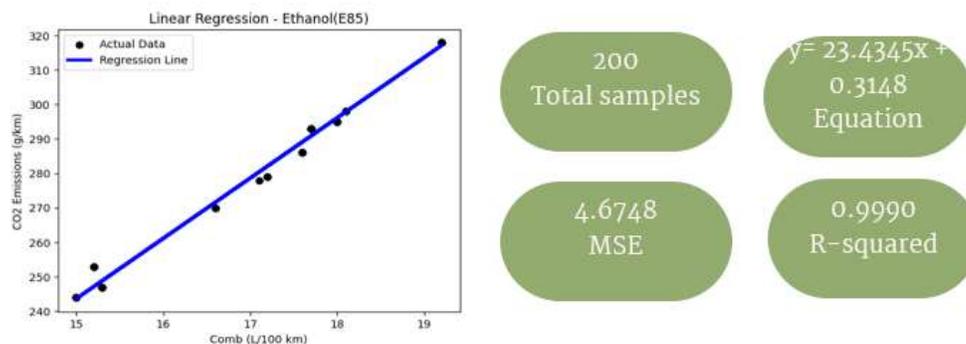
Berdasarkan gambar 2 diatas, terdapat 390 data sampel yang digunakan khusus untuk regular gasoline, dengan hasil persamaan regresi bernilai $y = 23.447x + 0.1701$ yang menunjukkan setiap kenaikan 1 unit, maka nilai akan bertambah sebesar 23.447. Kemudian untuk nilai MSE yang didapatkan sebesar 1.9844 yang artinya prediksi model mendekati nilai yang sebenarnya,

sedangkan untuk R – Squared bernilai 0.9994 menunjukkan bahwa hampir semua variabilitas dalam y dapat dijelaskan oleh variabel x, menandakan model regresi bekerja dengan sangat baik.



Gambar 3. Scatter plot untuk premium gasoline

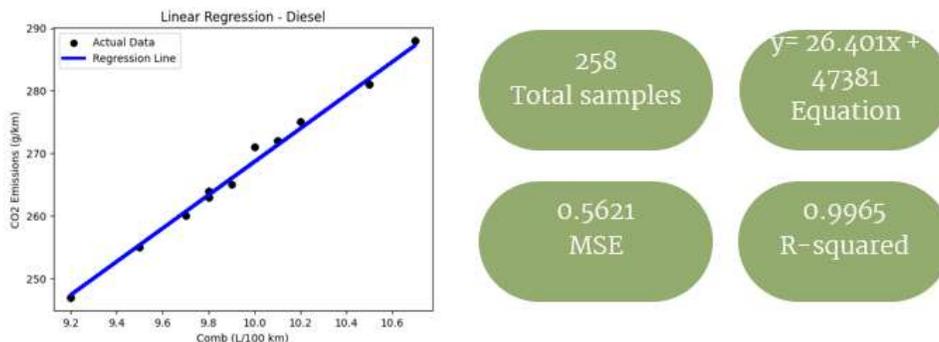
Berdasarkan gambar 3 diatas, terdapat 408 data sampel yang digunakan khusus untuk premium gasoline, dengan hasil persamaan regresi bernilai $y = 23.4345x + 0.3148$ yang menunjukkan setiap kenaikan 1 unit, maka nilai akan bertambah sebesar 23.4345. Kemudian untuk nilai MSE yang didapatkan sebesar 4.6748 yang artinya prediksi model cukup mendekati nilai yang sebenarnya, karena MSE menunjukkan besaran rata-rata kesalahan kuadrat, sedangkan untuk R – Squared bernilai 0.9990 menunjukkan bahwa hampir semua variabilitas dalam y dapat dijelaskan oleh variabel x, menandakan model regresi bekerja dengan baik



Gambar 4. Scatter plot untuk Ethanol

Berdasarkan gambar 4 diatas, terdapat 200 data sampel yang digunakan khusus untuk ethanol, dengan hasil persamaan regresi bernilai $y = 23.4345x + 0.3148$ yang menunjukkan setiap kenaikan 1 unit, maka nilai akan bertambah sebesar 23.4345. Kemudian untuk nilai MSE yang didapatkan sebesar 4.6748 yang artinya prediksi model cukup mendekati nilai yang sebenarnya, karena MSE menunjukkan besaran rata-rata kesalahan kuadrat, sedangkan untuk R – Squared bernilai 0.9990 menunjukkan bahwa hampir semua variabilitas dalam y dapat

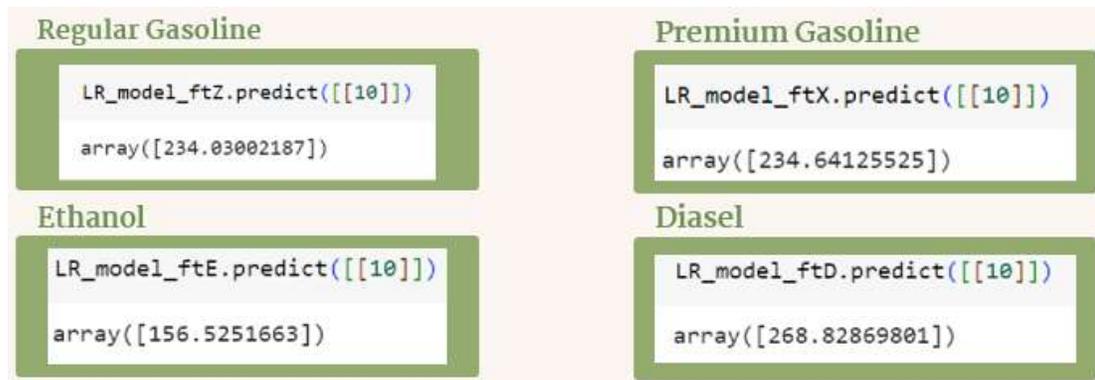
dijelaskan oleh variabel x, menandakan model regresi bekerja dengan sangat baik. Untuk hasil ethanol memiliki nilai yang persis dengan hasil premium gasoline.



Gambar 5. Scatter plot untuk Diesel

Berdasarkan gambar 5 diatas, terdapat 258 data sampel yang digunakan khusus untuk Diesel, dengan hasil persamaan regresi bernilai $y = 26.401x + 0.47338$ yang menunjukkan setiap kenaikan 1 unit, maka nilai akan bertambah sebesar 26.401. Kemudian untuk nilai MSE yang didapatkan sebesar 0.5621 yang artinya prediksi model sangat mendekati nilai yang sebenarnya dengan nilai yang hampir mendekati nol, sedangkan untuk R – Squared bernilai 0.9965 menunjukkan bahwa hampir semua variabilitas dalam y dapat dijelaskan oleh variabel x, menandakan model regresi bekerja dengan sangat baik. Untuk hasil regresi diesel bekerja dengan performa yang sangat baik.

Setelah menganalisis scatter plot dan performa dari model regresi linier yang telah dilatih sebelumnya menggunakan data test, maka model tersebut dapat digunakan untuk memprediksi data yang sebenarnya dengan inputan berupa nilai jarak tempuh kendaraan bermotor dan outputnya berupa emisi co2 yang dihasilkan. Berikut adalah penggunaan model berdasarkan 4 tipe bahan bakar yang digunakan:



Berdasarkan hasil dari gambar diatas, keempat model yang digunakan dapat memprediksi dengan baik, dimana apabila diberikan sebuah inputan nilai jarak tempuh sejauh 10 kilometer untuk bahan bakar yang digunakan, maka akan menghasilkan nilai emisi co2/gram. Untuk interface yang digunakan adalah google colaboratory untuk memudahkan dalam memproses data, hanya saja user interfacenya tidak ramah dengan pengguna oleh karena itu hasilnya dalam bentuk potongan kode.

KESIMPULAN DAN SARAN

Algoritma regresi linier yang diimplementasikan pada model bekerja dengan baik, hanya saja untuk setiap variabelnya memiliki hasil yang berbeda berdasarkan tipe bahan bakarnya. Hal ini dapat dipengaruhi oleh beberapa factor, diantaranya jumlah data yang diberikan pada saat pelatihan data dan persebaran data yang tidak merata, dapat dilihat pada scatter plot yang dihasilkan. Berdasarkan 4 tipe bahan bakar yang digunakan, tipe bahan bakar diesel memiliki performa yang paling baik, dengan nilai MSE sebesar 0.5621. Hasil emisi CO2 yang dihasilkan dihitung dengan satuan gram. Analisis ini hanya mencakup keseluruhan tipe kendaraan bermotor berdasarkan tipe bahan bakar yang digunakan, tidak menjelaskan secara spesifik jenis kendaraan yang digunakan sebagai subjek penelitian. Untuk saat ini model yang digunakan telah berhasil melakukan prediksi dengan performa yang cukup baik, diharapkan dengan model ini dapat dikembangkan model yang lebih kompleks dengan performa yang lebih baik dari model sebelumnya.

DAFTAR PUSTAKA

- A'yun, I.Q., Devi, A., (2023). Deteminants of CO2 Emission: Evidence from united States. *Journal of Economics Research and Social Sciences*. Vol 7, No 2.
- Dita, N.A., Kartika, E. S., Chairul, M. (2019). EMISI CO2 KENDARAAN BERMOTOR KAWASAN ALUN-ALUN KOTA BATU. *Planning for Urban Region and Environment* Volume 8, No 1.
- Guntara, G.R., (2023). Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Makser Wajah Menggunakan Algoritma Deep Learning YOLOv7. *Jurnal Teknologi dan Sistem Informasi Bisnis*. Vol 5, No 1.
- Harsiti., Zaenal, M., Ela, S., (2022). PENERAPAN METODE REGRESI LINIER SEDERHANA UNTUK PREDIKSI PERSEDIAAN OBAT JENIS TABLET. *Jurnal Sistem Informasi*. Vol 9. No 1. 12 – 16.
- Jainal, A., Erwina, A. H., Alwendi. (2022). Pentingnya Kesadaran Untuk Peduli Untuk Menjaga Dan Melestarikan Lingkungan. *Jurnal Pengabdian Masyarakat*. Vol 1. No. 3.
- Noviyanti, T. M. S., Fonggi, Y. A., (2022). Exploratory Data Analysis (EDA): A Study of Olympic Medallist. *SISTEMASI: Jurnal Sistem Informasi*. Volume 11, Nomor 3, 578-587.
- Sari, G.E., Sofwan, M., (2021). Carbon Dioxide (CO2) Emissions Due to Motor Vehicle Movement in Pekanbaru City Indonesia. *Journal of Geoscience Engineering, Environment, and Technology*. Volume 6, No.4.
- Soen, E.I.G., Marlina., Renny., (2022). Implementasi Cloud Computing dengan Google Colaboratory Pada Aplikasi Pengolah Data Zoom participants. *Jouurnal Informatic technology And Communication*. Vol 6, No. 1. pp 24-30.
- Theofani, G., Eko, s., (2022). Multiple Linear Regression Analysis on Factors that Influence Employees Work Motivation. *Jurnal dan Penelitian Teknik Informatika*. Volume 6, Nomor 3.
- Wahyuni, D.E., Amalia, A.A., Mashita, K. (2019). Explanatory Data Analysis dalam Konteks Klasifikasi Data Mining. *Prosiding Nasional Rekayasa Teknologi Industri dan Informasi XIV*. pp 263 – 269.