

PREDIKSI PENCAPAIAN INDEKS PRESTASI KUMULATIF (IPK) MENGGUNANA METODE LINEAR REGRESI

Imam Santoso^{1*}, Rahmat Nursiaga²

Program Studi Rekayasa Komputer^{1,2}

Fakultas Sains dan Teknologi^{1,2}

Universitas Teknologi Muhammadiyah Jakarta^{1,2}

*Corresponding author : imam.santoso@utmj.ac.id

Authors Email: imam.santoso@utmj.ac.id¹,
rahmat.nursiaga@utmj.ac.id²

Received: December 18,2025. **Revised:** February 7,2026. **Accepted:** February 10,2026. **Issue Period:** Vol.10 No.1 (2026), Pp. 297-303

Abstrak: Pendidikan tinggi sebagai salah satu lembaga bisnis yang bergerak di bidang layanan pendidikan tidak dapat dipisahkan dari jangkauan globalisasi. Perubahan dalam tren pendidikan dan pergerakan bebas ilmu pengetahuan dan teknologi yang merupakan aspek penting dari globalisasi akan menyentuh bidang pendidikan. Pengukuran akhir dari perkuliahan adalah dicapainya IPK sebagai pencapaian akhir proses perkuliahan. Untuk mengetahui prediksi nilai IPK sebelum semester akhir, maka dapat dilakukan dengan menggunakan algoritma *Multiple Linear Regression (MLR)* menggunakan variable independent Indek Prestasi Sementara (IPS) semester 1 dan 5. Dengan prediksi yang didapat pada akhir semester 5 diharapkan dapat membuat motivasi mahasiswa agar meningkatkan nilai di semester selanjutnya. Dengan dataset dari lulusan Universitas Teknologi Muhammadiyah Jakarta (UTM Jakarta) dibangun model prediksi dengan bahasa pemograman Python, menghasilkan model MLR yang moderat dengan nilai evaluasi MSE 0.016 dan R² 0.59 dengan interpretasi Model Baik dan akan lebih baik apabila ditambahkan ditambahkan variable independent lainnya.

Kata kunci: Prediksi IPK, *Multiple Linear Regression*, Python.

Abstract: Higher education as a business institution engaged in educational services cannot be separated from the reach of globalization. Changes in educational trends and the free movement of science and technology which are important aspects of globalization will affect the field of education. The final measurement of lectures is the achievement of the GPA as an achievement of the lecture process. To find out the predicted GPA value before the final semester, it can be done using the Multiple Linear Regression (MLR) algorithm using the Interim Grade Point Average (IPS) independent variable for semesters 1 and 5. With the prediction obtained at the end of semester 5, it is hoped that it can motivate students to improve their grades in the following semester. With a dataset from graduates of the Muhammadiyah University of Technology Jakarta (UTM Jakarta), a prediction model was built with the Python programming language, resulting in a moderate MLR model with an MSE evaluation value of 0.016 and R2 of 0.59 with the interpretation of a Good Model and will be better if other independent variables are added..

Keywords: *GPA Prediction, Multiple Linear Regression, Python.*



DOI: 10.52362/jisamar.v10i1.2304

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional.](https://creativecommons.org/licenses/by/4.0/)

I. PENDAHULUAN

Pendidikan tinggi sebagai salah satu lembaga bisnis yang bergerak di bidang layanan pendidikan tidak dapat dipisahkan dari jangkauan globalisasi. Perubahan dalam tren pendidikan dan pergerakan bebas ilmu pengetahuan dan teknologi yang merupakan aspek penting dari globalisasi akan menyentuh bidang pendidikan. Pendidikan tinggi adalah penyelenggara pendidikan akademik bagi mahasiswa. Mahasiswa sebagai produk pendidikan tinggi dapat digunakan sebagai referensi untuk menunjukkan keberhasilan pendidikan[1]. Prestasi belajar mahasiswa dapat dilihat berdasarkan Indeks Prestasi Kumulatif (IPK) yang diperoleh oleh mahasiswa. Maka pencapaian IPK mahasiswa bisa dijadikan dasar penilaian pencapaian hasil pendidikan dan pengajar pada perguruan tinggi.

Dalam proses akreditasi nilai rata-rata IPK mahasiswa juga salah satu yang dinilai. Karena IPK adalah bentuk evaluasi untuk proses pendidikan dan pengajaran. Tapi bukan berarti IPK adalah segalanya dalam pencapaian pembelajaran, karena masih ada faktor lain yang bisa dinilai sebagai keberhasilan dalam pendidikan dan pengajaran contohnya ada sikap atau sering disebut dengan *soft skill* [2]. IPK didapat dengan membagi hasil perkalian nilai mutu dan bobot kemudian dibagi dengan jumlah sks yang telah diambil pada seluruh semester. Maka besar jumlah sks yang telah diambil pada setiap semesternya akan sangat mempengaruhi besar kecil nilai IPK.

Disetiap semester mahasiswa akan mendapatkan Indeks Prestasi Semester/ Sementara (IPS) yang merupakan nilai evaluasi akhir dari proses pendidikan dan pengajaran pada semester yang berjalan. IPS dapat dijadikan gambaran untuk pencapaian IPK pada akhir pembelajaran. Karena IPK adalah gabungan IPS, mulai dari semester awal sampai dengan lulus. Penelitian ini akan memprediksikan nilai IPK yang didapat mahasiswa dengan memperhitungkan dari IPS pada 2 semester yaitu semester 1 dan semester 5 mahasiswa Universitas Teknologi Muhammadiyah Jakarta (UTM Jakarta) dengan menggunakan bahasa pemrograman python dengan algoritma *Multiple Linear Regression* (MLR).

Algoritma *Multiple Linear Regression* (MLR) dipilih karena dari sejumlah penelitian terdahulu telah menunjukkan efektivitas MLR dalam membangun model prediksi. Misal penelitian yang dilakukan oleh Zhao menekankan bahwa MLR tetap relevan sebagai metode klasik untuk forecasting, terutama dalam bidang energi dan kesehatan, serta dapat dioptimalkan dengan pendekatan komputasi modern[3]. Demikianpun dengan Dagdagui dalam penelitiannya menunjukkan bahwa MLR dapat digunakan untuk memprediksi performa akademik mahasiswa berdasarkan data demografis dan hasil ujian masuk, dengan hasil yang signifikan dalam mengidentifikasi faktor-faktor penentu prestasi[4]. Selain itu, penelitian lain juga menyoroti kemampuan MLR dalam memprediksi multivariat, yaitu ketika lebih dari satu variabel dependen dipengaruhi oleh sekumpulan variabel independen, yang memperluas cakupan aplikasinya dalam penelitian sosial maupun teknis [5].

II. METODE DAN MATERI

2.1. Bahasa Pemrograman Python

Python adalah bahasa pemrograman interpretatif serbaguna yang menggunakan filosofi desain yang berfokus pada keterbacaan kode. Sebagai bahasa pemrograman, Python menggabungkan kemampuan, sintaks, dan fungsi pustaka berkualitas tinggi[6] . Secara umum, bahasa pemrograman yang dirancang oleh Guido van Rossum ini banyak digunakan oleh masyarakat umum dalam pengembangan program. Misalnya, aplikasi smartphone, program CLI, program GUI (desktop), game, web, IoT, program peretasan, dan sebagainya[7]. Bahasa pemrograman ini menggunakan metode pemrograman modular, di mana pemrograman modular adalah teknik pemrograman yang memisahkan kode menjadi beberapa bagian yang disebut modul. Tujuannya adalah untuk mengurangi ketergantungan antar modul. Teknik ini biasanya digunakan untuk perangkat lunak berskala besar. Perangkat lunak ini memiliki banyak modul yang dibangun secara individual dan kemudian digabungkan untuk menjalankan fungsi perangkat lunak. Dalam Python, sebuah modul diimplementasikan dalam file .py yang berisi kumpulan kode untuk melakukan task tertentu. Sebuah modul dapat berisi *class*, *statement function* atau kode lainnya. Salah satu keuntungan menggunakan modul adalah kemampuan penggunaan kembali, artinya modul ini dapat dipanggil oleh berbagai modul lain.



DOI: 10.52362/jisamar.v10i1.2304

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](#).

2.2. Algoritma Linear Regression.

Regresi linear (linear regression) adalah metode yang digunakan untuk memperoleh model hubungan antara variabel dependen dan variabel independen. Terdapat dua jenis regresi linear, yaitu regresi linear sederhana (simple linear regression) dan regresi linear berganda (multiple linear regression)[8]. Perbedaannya terletak pada jumlah variabel independen. Regresi linear sederhana adalah metode yang digunakan untuk memperoleh model hubungan antara satu variabel dependen dan satu variabel independen, sedangkan regresi linear berganda adalah metode yang digunakan untuk memperoleh model hubungan antara satu variabel dependen dan lebih dari satu variabel independen. Analisis regresi sederhana berkaitan dengan hubungan antara dua variabel, khususnya variabel independen dan variabel dependen, sedangkan regresi ganda atau berganda melibatkan hubungan antara setidaknya tiga variabel, khususnya tentang dua variabel independen dengan satu variabel dependen.

a) Linear Regresi Berganda (*Multiple Linear Regression/ MLR*)

Multiple Linear Regression (MLR) merupakan salah satu metode analisis statistik yang digunakan untuk memodelkan hubungan antara variabel dependen dengan lebih dari satu variabel independen. Model ini sangat penting dalam penelitian kuantitatif karena mampu menjelaskan pengaruh simultan dari berbagai faktor terhadap suatu variabel target. Formula umum MLR dituliskan sebagai:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots + \beta_n x_n + \epsilon$$

di mana :

- Y : variabel dependen,
 $x_1, x_2 \dots x_n$: variabel independen,
 β_0 : merupakan intercept,
 $\beta_1, \beta_2 \dots \beta_n$: koefisien regresi,
 ϵ : error atau residual.

Efektivitas penggunaan MLR dalam memprediksi data ekonomi dengan akurasi tinggi, serta menunjukkan bagaimana integrasi dengan perangkat lunak Python dapat meningkatkan efisiensi analisis[9]. Dan MLR mampu mengidentifikasi variabel yang paling signifikan dalam memengaruhi performa akademik[10]. Dengan dukungan teori matematis yang kuat dan bukti empiris dari penelitian terdahulu, MLR terbukti sebagai metode yang relevan dan aplikatif. Kehadirannya dalam ekosistem Python melalui library seperti Scikit-learn semakin memperkuat posisinya sebagai alat prediksi yang efisien, mudah digunakan, dan dapat diandalkan dalam penelitian maupun aplikasi industri.

b) Skirt-Learn Library.

Skirt-Learn merupakan salah satu library Python yang paling banyak digunakan dalam bidang machine learning dan data mining. Library ini bersifat open-source dan menyediakan berbagai algoritma serta fungsi yang mendukung proses analisis data, mulai dari klasifikasi, regresi, clustering, hingga dimensionality reduction. Selain itu, Scikit-learn memiliki integrasi yang baik dengan library lain seperti NumPy, Pandas, dan Matplotlib, serta dilengkapi dengan dokumentasi yang komprehensif dan komunitas pengguna yang luas. Dalam konteks *Multiple Linear Regression* (MLR), Scikit-learn menyediakan kelas *LinearRegression()* yang memungkinkan pengguna untuk memodelkan hubungan antara variabel dependen dengan lebih dari satu variabel independen. Sejumlah penelitian terdahulu telah mendokumentasikan penggunaan Scikit-learn dalam regresi linear dapat digunakan untuk mengoptimalkan model regresi melalui pemilihan algoritma dan evaluasi performa[11]. Sementara itu penggunaan Skirt-Learn Library dibandingkan dengan beberapa library lainnya dalam implementasi MLR menunjukkan bahwa langkah-langkah pelatihan serta evaluasi model memiliki kesamaan yang signifikan [12].

2.3. Metode Penelitian.



DOI: 10.52362/jisamar.v10i1.2304

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](#).

Pada penelitian ini menggunakan teknik data mining metode forecasting/ prediksi, dengan memanfaatkan algoritma *Multiple Linear Regression* langkah-langkah metode penelitian dijelaskan pada gambar dibawah :



Data yang digunakan adalah data yang didapat dari Bagian Akademik UTM Jakarta mahasiswa yang telah lulus pada tahun 2024. Data yang didapat adalah sebanyak 120 mahasiswa, data yang digunakan hanya data mahasiswa yang masuk kuliah dari semester 1 sebagai mahasiswa baru pada 2 prodi yaitu S1 Manajemen dan S1 Akuntansi (mahasiswa pindahan atau konversi dari D3 ke S1 tidak digunakan). Fakultas yang diambil data lulusan hanya Fakultas Ekonomi dan Bisnis dikarenakan yang baru ada lulusan dari kampus ini adalah fakultas ini.

III. PEMBAHASA DAN HASIL

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini sebanyak 120 lulusan angkatan di wisuda pada wisuda ke III UTM Jakarta. Data disediakan oleh bagian akademik yang berisikan biodata lulusan dan data akademik, namun tidak semua data digunakan untuk penelitian ini. Yang digunakan adalah data Indek Prestasi Sementara semester 1 dan semester 5 serta .

3.2. Data Selection

Dari data set yang ada maka dilakukan data selection atau pemilihan data, karena tidak semua data yang tersedia dipakai dalam penelitian ini. Dari data yang diambil adalah Nilai Indek Prestasi Sementara semester 1 dan semester 5 serta Indek Prestasi Akumulatif (IPK) yang didapat saat kelulusan. Data ini hanya dari kelulusan mahasiswa prodi S1 Manajemen dan S1 Akuntansi, karena utk prodi pasca sarjana tidak bisa digunakan karena jumlah semester yang tidak ideal untuk dilakukan prediksi. Setelah itu data akan dilakukan pemilihan data untuk kelulusan mahasiswa yang mempunyai status masuk kuliah dengan mahasiswa baru, yang bukan merupakan mahasiswa pindahan ataupun konversi dari diploma menjadi sarjana.

3.3. Data Cleaning

Setelah dilakukan proses pemilihan data dan seleksi data selanjutnya melakukan proses pembersihan data. Pembersihan data dilakukan dengan cara membuang data duplikasi yang ada dan melengkapi data yang tidak lengkap.

3.4. Implementasi *Multiple Linear Regression* (MLR)

Dalam implementasi algoritma *Multiple Linear Regression* (MLR) dengan menggunakan bahasa pemrograman python digunakan dengan software code editor Ms. Visual Code. Adapun tahapan script code dijelaskan pada poin-poin berikut :

- Import Library yang digunakan program.

Script dibawah digunakan untuk import semua library yang dibutuhkan untuk menjalankan program, setelah itu program akan mengakses file data berbentuk csv dan kemudian menampilkan dataset sebagian dengan script *nilai_ipk.head()*



DOI: 10.52362/jisamar.v10i1.2304

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional.](https://creativecommons.org/licenses/by/4.0/)

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score

nilai_ipk = pd.read_csv("E:\\Dataset1.csv")
nilai_ipk.head()

```

Gambar 1. Script import library yang dibutuhkan dalam program

- b) Menampilkan deksripsi dataset.

Script `nilai_ipk.describe()` berfungsi untuk melakukan analisis deskriptif secara otomatis terhadap dataset yang dipilih dan menghasilkan output sesuai gambar 2.

- c) Script implementasi Algoritma MLR.

Pada script ini diawali dengan memilih variable yang digunakan untuk Variable Dependent (Y) dan Variable Independet (x_1 dan x_2). Kemudian script memisahkan dataset menjadi 2 bagian yakni menjadi train set dan test, dengan train setnya adalah 2/3 dari dataset yang ada. Kemudian pembuatan model regresi dengan script class fungsi `regressor = LinearRegression()`. Terakhir model akan melakukan prediksi berdasarkan model yang dibuat sebelumnya, script dapat diamati pada gambar 3.

nilai_ipk.describe()		
	IP51	IP55
count	40.000000	40.000000
mean	3.610750	3.480250
std	0.210581	0.289876
min	2.960000	2.890000
25%	3.500000	3.290000
50%	3.600000	3.570000
75%	3.725000	3.702500
max	4.000000	3.870000

Gambar 2. Dataset describe tentang dataset yang digunakan.

```

#Memilih Variable X dan y
X = nilai_ipk[['IP51','IP55']]
y = nilai_ipk['IPK']

#membagi data menjadi Data Training dan Data Test
X_train, X_test, y_train, y_test = train_test_split(X,y, test_size = 1/3, random_state=0)

#Fitting Simple Linear Regression Terhadap Training Set
regressor = LinearRegression()
regressor.fit(X_train, y_train)

#memprediksi hasil Test Set
y_pred = regressor.predict(X_test)

```

Gambar 3. Pembagian dataset (data train dan data test) dan pembuatan model *Multiple Linear Regression (MLR)*

- d) Evaluasi hasil prediksi dengan Residual Errors

Fungsi utama *residual error* (sisaan) dalam regresi adalah mengukur selisih antara nilai aktual dan nilai prediksi ($e = y - y'$) untuk mengevaluasi akurasi model. Residual digunakan untuk menentukan garis regresi terbaik (menggunakan metode kuadrat terkecil/*Ordinary Least Squares* - OLS), mendekripsi pola sistematis, serta memeriksa asumsi (normalitas, heteroskedastisitas) guna memastikan validitas model. Pada script ini memperlihatkan secara visual data hasil dari Residual Errors.



DOI: 10.52362/jisamar.v10i1.2304

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](#).

```
# plot for residual error
# setting plot style
plt.style.use('fivethirtyeight')

# plotting residual errors in training data
plt.scatter(regressor.predict(X_train),
            regressor.predict(X_train) - y_train,
            color='green', s=10,
            label='Train data')

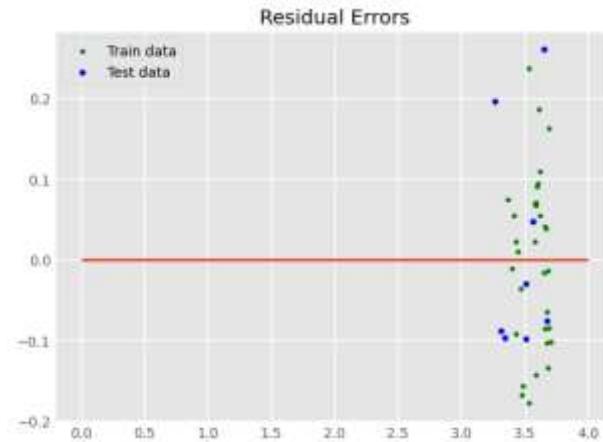
# plotting residual errors in test data
plt.scatter(regressor.predict(X_test),
            regressor.predict(X_test) - y_test,
            color='blue', s=10,
            label='Test data')

# plotting line for zero residual error
yfit = np.zeros(y.shape[0], dtype=y.dtype)

# plotting legend
plt.legend(loc='upper left')

# plot title
plt.title("Residual Errors")

# method call for showing the plot
plt.show()
```



Gambar 4. Visualisasi Data untuk Residual Errors

- e) Menampilkan nilai intercept dan kofisien dari model MLR.

Setelah model terbentuk maka untuk mengetahui model persamaan (intercept dan kofisien) dari MLR yang dihasilkan dapat menggunakan script pada gambar 5. Setelah di jalankan ternyata hasil yg ditampilkan adalah untuk intercept : 1.88, kofisien $x_1 = 0.088$ dan $x_2 = 0.39$. Maka dapat dituliskan sebagai berikut : $\textbf{Y} = \textbf{1.88} + \textbf{0.088}X_1 + \textbf{0.39}X_2$.

```
# regression coefficients
print('Coefficients: ', regressor.coef_)
print('Intercept: ', regressor.intercept_)
```

Gambar 5. Script menampilkan nilai kofisien dan intercept.

- f) Menampilkan evaluasi kekuatan Model.

Pada line script berikutnya (gambar 6) adalah pengukuran kekuatan model yang telah dibangun. Evaluasi yang digunakan adalah 2 metode yakni Mean Squared Error (MSE) dan R Squared (R^2). Dari script yang dibuat menampilkan nilai dari model adalah untuk $MSE = 0.016$ dan $R^2 = 0.59$. Artinya adalah :

- MSE 0.016 : model tergolong baik karena nilai MSE sangat kecil.
- $R^2 = 0.59$: model tergolong moderat, hasil akan lebih menghasilkan model yang robust bila ditambah dengan variable independen lainnya.

```
# Calculate the Mean Squared Error (MSE)
mse = mean_squared_error(y_test, y_pred)
# Calculate the R - squared (R2) score
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse}")
print(f"R - squared Score: {r2}")
```

Gambar 6. Script evaluasi model MLR

IV. KESIMPULAN

Berdasarkan pengujian menggunakan fungsi *Multiple Linear Regression (MLR)* untuk memprediksi Indek Prestasi Kumulatif/ IPK (Y) dari Indek Prestasi Sementara pada semester 1 dan 5 diketahui bahwa semakin bertambahnya nilai IPS maka makin bertambah nilai IPK yang akan didapat pada akhir perkuliahan. Walau dari hasil evaluasi model yang dibangun masih sangat perlu tambahan variable independen agar



DOI: 10.52362/jisamar.v10i1.2304

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional.](https://creativecommons.org/licenses/by/4.0/)

menambah robust nya model yang dibangun. Selain itu jumlah dataset yang digunakan bisa diperbanyak guna menemukan model prediksi yang lebih tepat.

REFERENASI

- [1] N. Almumtazah, N. Azizah, Y. L. Putri, and D. C. R. Novitasari, "Prediksi Jumlah Mahasiswa Baru Menggunakan Metode Regresi Linier Sederhana," *JURNAL ILMIAH MATEMATIKA DAN TERAPAN*, vol. 18, no. 1, pp. 31–40, Jun. 2021, doi: 10.22487/2540766x.2021.v18.i1.15465.
- [2] A. Irawan, T. Setiawati, and A. Andiana, "OPTIMALISASI SOFT SKILL PADA LEMBAGA PENDIDIKAN DI ERA 4.0," 2023.
- [3] Y. Zhao, S.-C. Chu, A. R. Yildiz, and J.-S. Pan, "Optimized Multiple Regression Prediction Strategies with Applications," *Symmetry (Basel)*., vol. 17, no. 7, p. 1085, Jul. 2025, doi: 10.3390/sym17071085.
- [4] R. T. Dagdagui, "Predicting Students' Academic Performance Using Regression Analysis," *Am. J. Educ. Res.*., vol. 10, no. 11, pp. 640–646, Nov. 2022, doi: 10.12691/education-10-11-2.
- [5] L. Breiman and J. H. Friedman, "Predicting Multivariate Responses in Multiple Linear Regression," *J. R. Stat. Soc. Series B Stat. Methodol.*., vol. 59, no. 1, pp. 3–54, Jan. 1997, doi: 10.1111/1467-9868.00054.
- [6] Rob Mastrodomenico, *The Python Book*. Jhon Wiley and Son Ltd, 2022.
- [7] D. Khulman, *A Python Book - Beginning Python, Advanced Python, and Python Exercises*. Platypus Global Media, 2014.
- [8] W. Ronald E, M. Raymond H, M. Sharon L, and Y. Keying, *Probability and Statistics For Engineers and Scientists Updated 9th Edition*. Pearson, 2017.
- [9] M. A. Shafi, M. Saifullah, S. Ismail, and M. Ghazali, "A Hybrid of Multiple Linear Regression Clustering Model with Support Vector Machine for Colorectal Cancer Tumor Size Prediction," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 4, 2019, doi: 10.14569/IJACSA.2019.0100439.
- [10] H. Midi, M. Sani, S. S. Ismaeel, and J. Arasan, "Fast improvised influential distance for the identification of influential observations in multiple linear regression," *Sains Malays.*., vol. 50, no. 7, pp. 2085–2094, Jul. 2021, doi: 10.17576/jsm-2021-5007-22.
- [11] M. Salama, "Optimization of Regression Models Using Machine Learning: A Comprehensive Study with Scikit-learn," *International Uni-Scientific Research Journal*, vol. 5, pp. 119–129, 2024, doi: 10.59271/s45500.024.0624.16.
- [12] A. Testas, "Multiple Linear Regression with Pandas, Scikit-Learn, and PySpark," in *Distributed Machine Learning with PySpark*, Berkeley, CA: Apress, 2023, pp. 53–74. doi: 10.1007/978-1-4842-9751-3_3.



DOI: 10.52362/jisamar.v10i1.2304

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional.](https://creativecommons.org/licenses/by/4.0/)