Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 2, February 2025, Page 553-562
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i2.6950

Prediksi Cuaca Menggunakan Data Historis dengan Algoritma Regresi Linear untuk Analisis Perubahan Suhu

Egi Pratama*, Muhammad Fatchan, Ahmad Aguswin

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia
Email: [1,*]egipratama.312110135@pelitabangsa.ac.id, [2]fatchan@pelitabangsa.ac.id, [3]aaguswin@pelitabangsa.ac.id
Email Penulis Korespondensi: egipratama.312110135@pelitabangsa.ac.id
Submitted: 08/02/2025; Accepted: 28/02/2025; Published: 28/02/2025

Abstrak−Tokyo, ibu kota Jepang yang terletak di pulau Honshu, menghadapi kompleksitas iklim subtropis, yang ditandai dengan variasi suhu ekstrem antara musim panas yang sangat panas (>35 derajat *Celsius*) dan musim dingin dengan suhu di bawah 0 derajat *Celsius*. Penelitian ini mengeksplorasi potensi algoritma regresi linear untuk memprediksi suhu maksimum harian dalam konteks dinamika cuaca perkotaan yang kompleks. Berdasarkan dataset meteorologi yang dikumpulkan selama total 639 hari, termasuk variabel-variabel kunci suhu, kelembaban, curah hujan, dan tekanan udara, penelitian ini mengembangkan model prediksi cuaca. Hasilnya menunjukkan kinerja yang luar biasa dengan *Root Mean Squared Error (RMSE)* sebesar 0,80 dan *R-squared* 0,99, yang menunjukkan kemampuan model untuk mencakup hampir seluruh pola variabilitas cuaca yang mungkin terjadi. Oleh karena itu, temuan penelitian tidak hanya mengonfirmasi efektivitas regresi linear untuk prediksi cuaca perkotaan, tetapi juga membuka kemungkinan integrasi model serupa dalam sistem prakiraan cuaca yang lebih canggih. Pendekatan yang berpusat pada data ini memberikan kontribusi signifikan pada teknologi prediksi cuaca modern yang responsif terhadap kebutuhan masyarakat perkotaan.

Kata Kunci: Prediksi Cuaca; Regresi Linear; Tokyo; Suhu Maksimum Harian; Data Meteorologi

Abstract−Tokyo, the capital of Japan located on the state of Honshu, is facing with subtropical climate complexity, combining extreme temperature variations reached in hot summer season (>35 degrees) and cold winter season temperatures below 0 degrees. Current research explored the regression linear algorithm potential to predict daily maximum temperature within the context of complex urban weather dynamics. Based on the meteorology dataset collected in total of 639 days including key variables of temperature, humidity, rainfall, and air pressure, study developed weather prediction model. The outcomes demonstrated exceptional performance with Root Mean Squared Error at 0.80 and R-squared of 0.99, showing the near full coverage of model's ability to capture all possible weather variability patterns. As a result, the research findings not only confirmed the effectiveness of linear regression for urban weather prediction but also open the possibility of similar model integration within more sophisticated weather forecast systems. Data-centered approach made significant contribution to the modern weather prediction technology responsive to urban society requirement.

Keywords: Weather Prediction; Linear Regression; Tokyo; Daily Maximum Temperature; Meteorological Data

1. PENDAHULUAN

Prediksi cuaca di Tokyo penting untuk menilai risiko kesehatan di masa depan yang disebabkan oleh perubahan iklim dan masyarakat yang menua. Teknik pembelajaran mesin telah digunakan untuk memprediksi kematian akibat penyakit kardiovaskular seperti infark miokard dan infark serebral di Tokyo selama bulan-bulan musim panas. *Transfer learning* telah diterapkan untuk mengevaluasi efek perubahan iklim terhadap risiko kematian di Tokyo. Teknik *over-sampling* telah disarankan untuk penelitian mendatang[1]. Dengan memahami hubungan antara cuaca dan hasil kesehatan, pembuat kebijakan dapat menerapkan intervensi yang ditargetkan untuk melindungi populasi yang rentan dengan kemajuan teknologi dan analisis data yang terus berlanjut, kita semakin dekat untuk menciptakan masyarakat yang lebih sehat dan lebih tangguh dalam menghadapi perubahan iklim. Seiring dengan terus mengumpulkan lebih banyak data dan menyempurnakan model, penelitian ini dapat lebih baik mengantisipasi dan mempersiapkan dampak kesehatan dari peristiwa cuaca ekstrem, seperti gelombang panas dan bencana alam. Pendekatan proaktif ini pada akhirnya dapat mengurangi beban pada sistem kesehatan dan meningkatkan kesejahteraan masyarakat secara keseluruhan[2].

Penelitian serupa mengenai model prediksi curah hujan dengan metode regresi linear banyak dilakukan. Dengan menggunakan regresi linear berganda akan mengahasilkan output yang relatif lebih baik jika dibandingkan dengan satu parameter cuaca sebagai *prediktor*[3].

Prediksi adalah proses membuat perkiraan atau ramalan tentang suatu kejadian atau hasil di masa depan berdasarkan informasi atau data yang tersedia saat ini. Tujuan dari prediksi adalah untuk memperkirakan kemungkinan hasil atau kejadian di masa depan, sehingga dapat membantu pengambilan keputusan yang lebih baik dan lebih tepat. Prediksi dapat dilakukan dengan berbagai cara, termasuk penggunaan statistik, analisis data, atau teknik kecerdasan buatan seperti *machine learning*. Dalam banyak kasus, prediksi melibatkan penggunaan model matematika atau statistik yang dibuat berdasarkan data historis atau informasi lain yang relevan[4].

Regresi linier adalah teknik analisis statistik yang digunakan untuk menemukan hubungan fungsional antara dua variabel, di mana satu variabel (variabel independen) mempengaruhi atau memprediksi nilai variabel lainnya (variabel dependen). Tujuan regresi linier adalah untuk menemukan garis terbaik (*best fit line*) yang dapat digunakan untuk memprediksi nilai variabel dependen berdasarkan nilai variabel independen yang diberikan. Regresi linier dapat digunakan dalam berbagai bidang, termasuk ilmu sosial, ekonomi, ilmu

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 2, February 2025, Page 553-562
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i2.6950

lingkungan, dan ilmu alam. Misalnya, regresi linier dapat digunakan untuk memprediksi penjualan produk berdasarkan biaya iklan, atau untuk menentukan hubungan antara suhu, penyakit dan kepadatan air di lingkungan alam[4].

Adapun beberapa penelitian sebelumnya yang telah memberikan dukungan bagi penelitian ini. Penelitian pertama yang dilakukan oleh Ardytha Luthfiarta, dkk dengan judul "Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda" pada tahun 2020. Penelitian ini dikembangkan melalui pengolahan data sekunder *database* kesehatan Dataset Diabetes yang diambil dari dataset BMKG dari Stasiun Meteorologi Ahmad Yani Semarang tahun 2015-2017. Kemudian, mengolah data diabetes dengan akurasi akan dievaluasi dengan menggunakan algoritma Regresi Linear Berganda[5]. Penelitian kedua yang dilakukan oleh Miftahuljannah dkk yang berjudul "Analisis Prediksi Penjualan Dengan Metode Regresi Linear Di Pt. *Eagle Industry* Indonesia" pada tahun 2023. Penelitian ini adalah untuk memprediksi perencanaan penjualan produk sehingga perencanaan dalam mengambil keputusan akan lebih mudah. Dengan data, penelitian ini memprediksi perencanaan penjualan dengan menggunakan data berupa produk, *plan*, dan *actual*. Dengan hasil evaluasi *Root Mean Squared Error* adalah 36241.241 +/- 0.000, dan *Squared Error* adalah 1313427569.481 +/- 5882150128.134. Hal Ini menunjukkan bahwa *Squared Error* menghasilkan nilai yang tinggi dari Root Mean Squared Error[6]. Penelitian ketiga yang dilakukan oleh Dini Rizki Septiani dkk yang berjudul "Pengembangan Model Prediksi Cuaca Menggunakan Teknik *Machine Learning*" pada tahun 2024. Penelitian ini bertujuan untuk mengembangkan model prediksi cuaca menggunakan teknik *machine learning* untuk meningkatkan akurasi dan ketepatan waktu dalam ramalan cuaca. Prediksi cuaca memiliki peran penting dalam berbagai sektor, mulai dari pertanian hingga transportasi, namun, prediksi yang akurat dan tepat waktu masih menjadi tantangan yang kompleks dalam ilmu meteorologi. Metode tradisional sering kali mengandalkan model fisika yang rumit dan memerlukan waktu komputasi yang lama, sementara itu, pendekatan *machine learning* menawarkan alternatif yang lebih cepat dan efisien. Dalam penelitian ini, data cuaca historis dari berbagai sumber seperti stasiun cuaca, satelit, dan sensor udara dikumpulkan dan diproses untuk melatih model *machine learning*[7].

Penelitian keempat yang dilakukan oleh Edi dkk yang berjudul "Prediksi Harga pada *Trading Forex Pair USDCHF* Menggunakan Regresi Linear" pada tahun 2023. Tujuan penelitian ini adalah membuat model prediksi harga *forex* untuk mempermudah *trader* melakukan prediksi harga. Dataset sebanyak 2066 data diperoleh melalui *software metatrader* dan diproses melalui tahap *preprocessing.* Model regresi linear dibuat menggunakan 5 skenario dan evaluasi dilakukan menggunakan nilai *Mean Squared Error (MSE)* dan *Root Mean Square Error (RMSE)* untuk memilih model terbaik. Hasilnya menunjukkan bahwa regresi linear mampu memprediksi harga penutupan pada *pair USDCHF*. Model regresi linear terbaik diperoleh menggunakan variabel bebas pada skenario 1, yaitu variabel *Open*, dengan persamaan regresi linear y=0,0145+0,9849x, *MSE* terbaik sebesar 0,0000328509 dan RMSE terbaik sebesar 0,0057315705[8]. Penelitian kelima yang dilakukan oleh Andrianto dan Irawan yang berjudul "Implementasi Metode Regresi Linear Berganda Pada Sistem Prediksi Jumlah Tonase Kelapa Sawit di PT. Paluta Inti Sawit" pada tahun 2023. Kabupaten Padang Lawas Utara memiliki luas perkebunan kelapa sawit tahun 2021 yaitu seluas 27.776 Ha (40%). Kecamatan Halongonan menjadi sentral pertumbuhan dan perkembangan utama perkebunan kelapa sawit di Kabupaten Padang Lawas Utara seluas 6.477 Ha (23,55%) di tahun 2021. Sistem prediksi jumlah tonase buah kelapa sawit menggunakan metode regresi linear berganda. Secara keseluruhan tingkat akurasi prediksi jumlah tonase kelapa sawit selama satu bulan sebesar 99,99%, tingkat akurasi prediksi terendah pada tanggal 03 november 2022 sebesar 88%, sedangkan akurasi prediksi tertinggi pada tanggal 05 november 2022 dan tanggal 07 november 2022 sebesar 100%[9]. Penelitian keenam yang dilakukan Halawa dkk yang berjudul "*Implementation of Linear Regression Algorithm to Predict Stock Prices Based on Historical Data"* pada tahun 2022. Algoritma regresi linear adalah metode statistik yang digunakan untuk menentukan pengaruh satu atau beberapa variabel terhadap suatu variabel lainnya. Variabel merupakan besaran yang nilainya dapat berubah. Variabel yang memengaruhi disebut sebagai variabel independen, variabel bebas, atau variabel penjelas,variabel yang dipengaruhi disebut sebagai variabel dependen atau variabel terikat. Dalam analisis regresi, terdapat dua jenis variabel utama, yaitu variabel dependen adalah variabel yang keberadaannya dipengaruhi oleh variabel lain. Variabel ini tidak dapat berdiri sendiri dan dilambangkan dengan Y. Variabel independen adalah variabel yang tidak dipengaruhi oleh variabel lain, memiliki sifat mandiri, dan dilambangkan dengan X[10]. Penelitian ketujuh yang dilakukan Bahtiar dkk pada tahun 2023 mengenai penerapan metode Regresi Linier, yakni metode yang digunakan untuk menguji hubungan antara suatu variabel tersier dengan dua atau lebih variabel sekunder. Berdasarkan prediksi menggunakan bahasa pemrograman Python hasil panen padi tahun 2023 sebanyak 1510403 Ton/GKP, dengan nilai *MAE, MSE, RMSE, R2-Score*, dan sistem menampilkan nilai *MAE (Mean Absolute Error)* : 5449.45, nilai *MSE (Mean Squaed Error)*: 72325540.80, *RMSE (Roots Mean Squaed Error)* : 8504.44,dan *R2-Score* : 0.93 dengan prediksi di tahun 2023 mengalami penurunan dari tahun sebelumnya[11]. Penelitian kedelapan yang di lakukan Wijaya dkk yang berjudul "Penerapan Data Mining Pada Prediksi Harga Emas dengan Menggunakan Algoritma Regresi Linear Berganda dan ARIMA" pada tahun 2023. Pada penelitian ini akan dilakukan proses penelitian dengan melakukan perbandingan dari algoritma Regresi Linear Berganda. Perbandingan dari algoritma bertujuan untuk mendapatkan hasil yang paling optimal dari penerapan algoritma. Dalam penyelesaian dengan menggunakan algoritma Regresi Linear Berganda dan ARIMA kedua algoritma tersebut dapat membantu menyelesaikan

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 2, February 2025, Page 553-562
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i2.6950

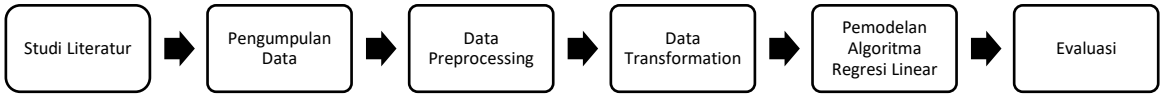
permasalah prediksi dengan menghasilkan hasil yang optimal. Dari proses yang dilakukan algoritma Regresi Linear Berganda memiliki nilai RMSE sebesar 4902782.346 sedangkan untuk algoritma ARIMA mendapatkan nilai sebesar 5876287.332. Hal tersebut menandaka bahwasannya hasil dari algoritma Regresi Linear Berganda lebih baik dibandingkan dengan algoritma ARIMA[12]. Penelitian kesembilan yang dilakukan Pebrilia yang berjudul "JIFP (Jurnal Ilmu Fisika dan Pembelajarannya) Analisis Curah Hujan Menggunakan *Machine Learning* Metode Regresi Linier Berganda Berbasis *Python dan Jupyter Notebook Rainfall Analysis using Machine Learning-Multiple Linear Regression Method Based on Python and Jupyter Notebook*" pada tahun 2022. Pada penelitian ini, telah dilakukan prediksi curah hujan dengan melibatkan tiga variable bebas yaitu kecepatan angin, suhu udara maksimum,dan suhu udara minimum dengan dataset diperoleh dari situs *kaggle.com*. Dataset yang digunakan berjumlah 6.574 data, dimana data tersebut dikelompokkan ke dalam data *training* sebanyak 80% dan data *test* sebanyak 20%. Algoritma regresi linier berganda dibuat dalam Bahasa pemrograman *python* dan diimplementasikan menggunakan *jupyter notebook*. Pada penelitian ini dihasilkan model regresi linier berganda dengan persamaan $y = 1.23 + 0.1x\,1 - 0.06x\,2 + 0.07x\,3$, nilai *MSE* sebesar 14.02, *RMSE* sebesar 3.74, dan *MAE* sebesar 2.27[13]. Penelitian kesepuluh yang dilakukan Alifi dkk yang berjudul "Penerapan Algoritma Regresi Linier pada Prediksi Tarif *Influencer Media Sosial*" pada tahun 2022. Penelitian ini bertujuan untuk memberikan solusi berupa model prediksi tarif i*nfluencer* berbasis *machine learning* yang dapat dijadikan referensi untuk meminimalisir dampak kerugian baik bagi *influencer* dalam menawarkan tarif, maupun *klien* dalam menerima tawaran tarif. Tahapan penelitian ini terdiri dari studi pustaka, pengumpulan data, pra-pemrosesan data, pembangunan model regresi linier, dan evaluasi model. Penelitian ini menghasilkan lima varian model. Salah satu model terbaik menghasilkan nilai *MAE*: 145401.484375, *MSE*: 7.222241e+10 dan *RMSE*: 268742.250, yang dipengaruhi oleh nilai *hyperparameter learning rate:* 0.001 dan *epoch:* 1.000. Model tersebut belum sepenuhnya mampu mewakili data mayoritas yang diujicoba[14].

Berdasarkan berbagai penelitian sebelumnya yang telah membuktikan efektivitas regresi linear dalam prediksi cuaca dan fenomena lainnya, penelitian ini berkontribusi dengan fokus spesifik pada prediksi suhu maksimum harian di Tokyo menggunakan data historis. Tidak hanya mengevaluasi performa algoritma regresi linear dalam konteks iklim perkotaan yang kompleks, penelitian ini juga memberikan wawasan baru dalam memahami hubungan antara variabel meteorologi utama dan perubahan suhu ekstrem. Dengan pendekatan berbasis data dan pemodelan statistik yang akurat, hasil penelitian ini diharapkan dapat meningkatkan efisiensi sistem prakiraan cuaca, memberikan informasi yang lebih akurat bagi pembuat kebijakan, serta memperkuat kesiapsiagaan masyarakat dalam menghadapi dampak perubahan iklim. Oleh karena itu, penelitian ini tidak hanya melengkapi literatur ilmiah yang ada, tetapi juga membuka peluang bagi pengembangan model prediksi cuaca yang lebih presisi di masa mendatang.

2. METODOLOGI PENELITIAN

2.1 Kerangka Penelitian

Dalam penelitian ini, penulis melakukan beberapa tahapan. Tahap-tahap tersebut adalah Studi literatur, pengumpulan data, data preprocessing, data tranformation, pembuatan model dan terakhir evaluasi data. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Penelitian

Pada Gambar 1, dapat dilihat bahwa tahap yang pertama yaitu studi literatur, proses ini dilakukan dengan meninjau beberapa jurnal terdahulu yang digunakan sebagai referensi pada penelitian ini. Setelah itu tahap kedua yaitu proses pengumpulan data. Pengumpulan data dilakukan dengan mencari data sekunder yaitu dataset historis cuaca di Tokyo yang terdapat pada *website visual crossing weather*. Tahap ketiga yaitu *data preprocesing* yang dilakukan dengan menormalisasi data agar dapat diolah lebih mudah dan tertata. Tahap keempat yaitu data transformation yaitu data yang akan dinormalisasi, Normalisasi dilakukan agar tiap fitur pada dataset memiliki rentang yang sama yaitu 0 sampai 1 menggunakan *MinMaxScaler* agar tidak memiliki ketimpangan rentang fitur. Tahap selanjutnya adalah pemodelan model algoritma Regresi Linear metode statistik yang luas digunakan untuk memodelkan hubungan antara variabel independen dan variabel dependen dalam bentuk persamaan garis lurus. Dan dievaluasi menggunakan metrik seperti seperti *Mean Squared Error(MSE),Root Mean Squared Error(RMSE)* dan Koefisien Determinasi (R2) untuk mengukur seberapa baik model dapat memprediksi nilai variabel dependen.

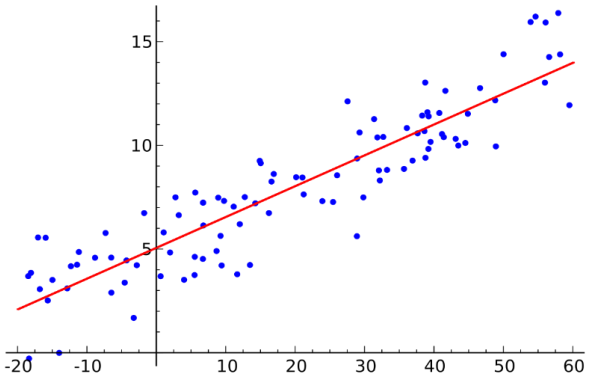
2.2 Regresi Linear

Regresi linear adalah metode statistik yang luas digunakan untuk memodelkan hubungan antara variabel independen dan variabel dependen dalam bentuk persamaan garis lurus. Namun, tidak seperti model klasifikasi,

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 2, February 2025, Page 553-562
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i2.6950

regresi linear tidak sesuai untuk memprediksi hasil kategorikal, dan oleh karena itu, *confusion matriks* (TP, TN, FP, FN) tidak langsung dapat diterapkan. Sebaliknya, kinerja model regresi linear biasanya dievaluasi menggunakan metrik seperti *Mean Squared Error(MSE),Root Mean Squared Error (RMSE)* dan Koefisien Determinasi (R2) untuk mengukur seberapa baik model dapat memprediksi nilai variabel dependen[15].
Proses kerja diatur sebagai berikut.

1. Mengukur rata-rata kuadrat perbedaan antara nilai prediksi dan nilai aktual
2. Mengukur proporsi variansi dalam variabel dependen yang dapat diprediksi dari variabel independen. Ini adalah ukuran kecocokan model, dengan nilai yang lebih tinggi menunjukkan kinerja yang lebih baik.



Gambar 2. Linear Regresi

2.3 *Mean Squared Error (MSE)*

MSE mengukur rata-rata kuadrat perbedaan antara nilai prediksi dan nilai aktual. Ini adalah ukuran akurasi model, dengan nilai yang lebih rendah menunjukkan kinerja yang lebih baik[16].
Rumus :

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y})^2 \quad (1)$$

Keterangan dari rumus tersebut sebagai berikut : n adalah jumlah data, yi adalah nilai aktual ke-i, ỹi adalah nilai prediksi ke-i.

2.4 *Root Mean Squared Error (RMSE)*

Metode ini digunakan untuk mengetahui besarnya penyimpangan yang terjadi antara nilai prediksi total curah hujan dibandingkan dengan nilai total curah hujan aktualnya yang terjadi selama satu tahun. Perlu diketahui bahwa untuk verifikasi hasil prakiraan semakin besar nilai RMSE, maka semakin jauh nilai prediksi total curah hujan bulanan terhadap nilai aktualnya dan semakin kecil nilai RMSE maka semakin baik prediksi total curah hujan bulanannya. Hal ini karena tingkat kesalahan yang dapat diminimalisasi dapat meningkatkan tingkat akurasi prediksi[17].
Rumus :

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y})^2} \quad (2)$$

Keterangan dari rumus tersebut sebagai berikut : n adalah jumlah observasi, yi adalah nilai aktual, ŷ adalah nilai prediksi.

2.5 Koefisien Determinasi (R2)

R2 mengukur proporsi variansi dalam variabel dependen yang dapat diprediksi dari variabel independen. Ini adalah ukuran kecocokan model, dengan nilai yang lebih tinggi menunjukkan kinerja yang lebih baik[18].
Rumus :

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (3)$$

Keterangan dari rumus tersebut sebagai berikut : R2: Koefisien determinasi, yang menunjukkan proporsi varians variabel dependen yang dijelaskan oleh variabel independen, 1: Menunjukkan bahwa seluruh varians variabel dependen akan dijelaskan oleh model jika R2 sama dengan 1, SSres: Jumlah kuadrat residual, yaitu jumlah perbedaan kuadrat antara nilai aktual dan nilai yang diprediksi, SStot: Jumlah kuadrat total, yaitu jumlah perbedaan kuadrat antara nilai aktual dan rata-rata nilai aktual.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 2, February 2025, Page 553-562
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i2.6950

2.6 Dataset

Dataset (Data set) adalah sekumpulan data atau himpunan data. Entitas *(entity)* diwakili oleh kumpulan data, yang merupakan kumpulan data yang menjelaskan item dengan properti objek. Dalam arti teknis, data adalah kumpulan nilai variabel kualitatif atau kuantitatif tentang satu orang atau sekelompok individu. Pemrosesan data seringkali bertahap. Data diukur, dikumpulkan, dilaporkan, dianalisis, dan digunakan untuk menghasilkan visualisasi data termasuk gambar, peta, tabel, grafik, dan infografis[19]. Data set yang digunakan adalah dataset publik dari *Visual Crossing Weather Data*, diperoleh dari *website Visual Crossing* yaitu: https://www.visualcrossing.com/weather/weather-dataservices/tokyo/metric/last15days. Dataset Awal Cuaca yang diperoleh dari *website Visual Crossing Weather Data.* Dataset cuaca terdiri dari 639 data dan 6 atribut yang terdiri dari suhu minimum,suhu,kelembaban,kecepatan angin,curah hujan,suhu maksimum. Atribut cuaca ditunjukan pada Gambar 3.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 639 entries, 0 to 638
Data columns (total 6 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   tempmin    639 non-null    float64
 1   temp       639 non-null    float64
 2   humidity   639 non-null    float64
 3   windspeed  639 non-null    float64
 4   precip     639 non-null    float64
 5   tempmax    639 non-null    float64
dtypes: float64(6)
memory usage: 30.1 KB
```

Gambar 3. Atribut yang digunakan Prediksi Cuaca di Tokyo

3. HASIL DAN PEMBAHASAN

3.1 *Data Preprocessing*

Data Preprocessing merupakan salah satu tahapan dalam melakukan mining data. Sebelum menuju ke tahap pemprosesan. Data mentah akan diolah terlebih dahulu. *Data Preprocessing* atau praproses data biasanya dilakukan melalui cara eliminasi data yang tidak sesuai. Selain itu dalam proses ini data akan diubah dalam bentuk yang akan lebih dipahami oleh sistem[20]. Pada tahap ini sebelum masuk ketahap modeling dataset terlebih dahulu diproses melalui tahap *data cleaning, Handling outlier* dan *data transformation*.

1. *Data Cleaning*
   Pada proses ini dilakukan pengecekan missing value, memastikan data tidak ada yang hilang. Hasil dari eksekusi bahwa dataset ini sudah tidak ditemukan adanya *missing value* ditunjukan pada Gambar 4

```
df.isnull().sum()
```

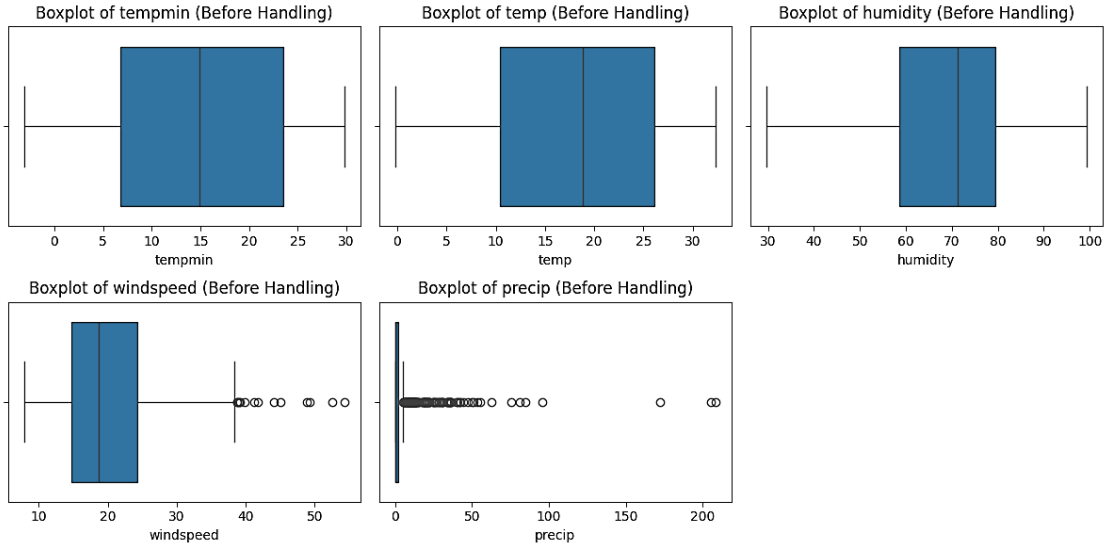
	0
tempmin	0
temp	0
humidity	0
windspeed	0
precip	0
tempmax	0

dtype: int64

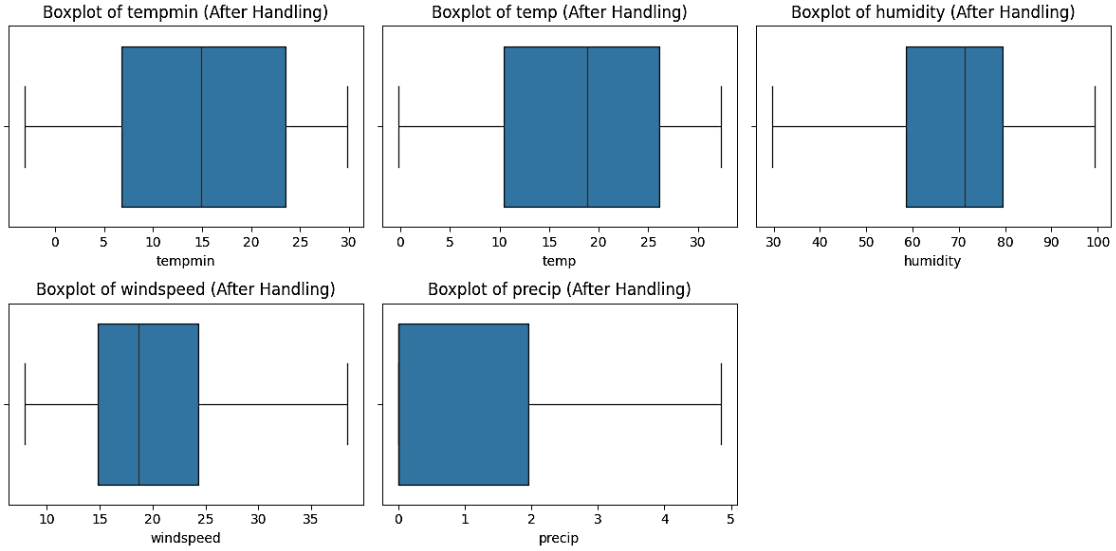
Gambar 4. Check Missing Value

Handling Outlier, Salah satu metode untuk handling outlier adalah menggunakan *IQR (Interquartile Range)*. Pada Gambar 5 menunjukan bahwa masing-masing atribut memiliki outlier dan atribut insulin yang paling banyak memiliki *outlier*. Oleh karena itu perlu dilakukan pengisian *outlier* dengan bantuan nilai IQR. Semua titik data yang memiliki nilai kurang dari $(1{,}5 \cdot IQR)$ di bawah kuartil pertama atau lebih dari $(1{,}5 \cdot IQR)$ di atas kuartil ketiga akan dilakukan imputasi fitur dengan nilai mediannya.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 2, February 2025, Page 553-562
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i2.6950



Gambar 5.Sebelum Handling Outlier



Gambar 6. Sesudah Handling Outlier

2. *Data Transformation*
Selanjutnya transformasi data adalah data yang akan dinormalisasi. Normalisasi dilakukan agar tiap fitur pada dataset memiliki rentang yang sama yaitu 0 sampai 1 menggunakan *MinMaxScaler* agar tidak memiliki ketimpangan rentang fitur. Pada Gambar 7 dapat dilihat data yang sudah ternormalisasi. Data yang sudah di transformasikan menggunakan rumus *MinMax Scaling*, sebagai berikut:

```
# Data Transformation
# Feature Scaling (MinMaxScaler)
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
# Fit and transform the features
scaled_features = scaler.fit_transform(df[features])
# Create a new DataFrame with scaled features
df_scaled = pd.DataFrame(scaled_features, columns=features, index = df.index)
df_scaled['tempmax'] = df['tempmax']

print(df_scaled.head())
```

```
    tempmin      temp  humidity  windspeed  precip  tempmax
0  0.133739  0.206154  0.473458   0.108197     0.0     12.6
1  0.173252  0.200000  0.311334   0.154098     0.0     11.3
2  0.121581  0.184615  0.176471   0.167213     0.0     10.6
3  0.142857  0.178462  0.255380   0.219672     0.0     10.3
4  0.188450  0.190769  0.087518   0.406557     0.0     10.0
```

Gambar 7. Transformasi Data

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 2, February 2025, Page 553-562
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i2.6950

2.2 *Data Modeling*

Menghitung Regresi Linear menggunakan *Colab*. Dalam analisis data ini, kita memodelkan hubungan antara beberapa variabel independen dan variabel dependen. Kode berikut menggunakan pustaka *numpy* untuk manipulasi *array* dan *statsmodels* untuk analisis regresi. Data yang digunakan mencakup 5 variabel independen: *Tempmin, Temp, Humidity, Windspeed,* dan *Precip*. Kemudian 1 variabel dependen yaitu *Tempmax.* Tujuan dari model ini adalah untuk menentukan bagaimana variabel-variabel independen mempengaruhi variabel dependen untuk menentukan suhu maksimal. Selanjutnya membagi data training dan data testing. Tujuan dari pemisahan data pelatihan dan pengujian adalah untuk menguji kinerja model pada data yang belum pernah dilihat sebelumnya dan untuk memastikan bahwa model tersebut tidak hanya cocok dengan set pelatihan. Pada tahap *Data splitting,* dilakukan pemembagian data menjadi 80% *Data Training* dan 20% *Data Testing* .

```
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
print(X.shape, X_train.shape, X_test.shape)
print(y.shape, y_train.shape, y_test.shape)
```

```
(639, 5) (511, 5) (128, 5)
(639,) (511,) (128,)
```

Gambar 8. Splitting Data Train & Data Test

```
# Build the Model
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(X_train, y_train)
```

```
LinearRegression
LinearRegression()
```

Gambar 9. Import Library

3.2 Pengujian

Pada pengujian model Regresi Linear, data dibagi menjadi data latih dan data uji. Model dilatih menggunakan data latih, kemudian diuji performanya menggunakan data uji. Evaluasi model dilakukan dengan metrik *RMSE (Root Mean Squared Error)* dan *R-squared.* RMSE mengukur seberapa dekat prediksi model dengan nilai sebenarnya, semakin kecil nilai RMSE semakin baik. R-squared mengukur seberapa baik model menjelaskan variasi data, semakin dekat nilai *R-squared* ke 1 semakin baik. Dari pengujian ini, diharapkan model Regresi Linear yang dihasilkan memiliki nilai RMSE yang rendah dan R-squared yang tinggi, menandakan model yang akurat dan handal dalam memprediksi data.

```
# Lakukan prediksi
y_pred = model.predict(X_test)

# Hitung RMSE
rmse = np.sqrt(mean_squared_error(y_test, y_pred))

# Hitung R-squared
r2 = r2_score(y_test, y_pred)

print(f"RMSE: {rmse:.2f}")
print(f"R-squared: {r2:.2f}")
```

```
RMSE: 0.80
R-squared: 0.99
```

Gambar 10. Pengujian Nilai RMSE & R-Squared

Dengan hasil pengujian ini menunjukkan nilai *RMSE* 0.80 dan *R-squared* 0.99 (99%) dengan ini model regresi yang diuji memiliki kinerja yang sangat baik dalam memprediksi data. *R-squared* yang sangat tinggi menunjukkan bahwa model memiliki kemampuan yang baik dalam menjelaskan variasi data, dan RMSE yang rendah menunjukkan kesalahan prediksi yang kecil.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 2, February 2025, Page 553-562
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i2.6950

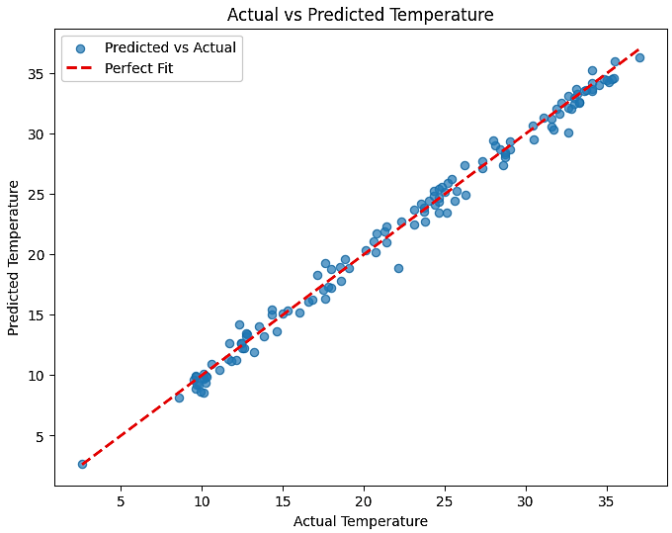
```
#Step 6: Visualize the Results
plt.figure(figsize=(8, 6))
plt.scatter(y_test, y_pred, alpha=0.7, label="Predicted vs Actual")
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], '--r', linewidth=2, label="Perfect Fit")
plt.xlabel("Actual Temperature")
plt.ylabel("Predicted Temperature")
plt.title("Actual vs Predicted Temperature")
plt.legend()
plt.show()
```

Gambar 11. Sintaks Hasil Visualisasi

Sintaks diatas merupakan kode *Python* yang menggunakan *library matplotlib.pyplot* untuk memvisualisasikan hasil dari model Regresi Linear dalam bentuk grafik *scatter plot*. Dengan ini, kita dapat melihat secara visual seberapa baik model Regresi Linear dalam memprediksi nilai suhu dibandingkan dengan nilai suhu aktual. Titik-titik yang mendekati garis *"perfect fit"* menunjukkan prediksi yang akurat, sementara titik-titik yang jauh dari garis menunjukkan kesalahan prediksi yang lebih besar.

3.3 Hasil Penelitian

Hasil penelitian dari model prediksi yang dilakukan dengan Algoritma Regresi Linear sebagai berikut. Dari hasil pengujian ini menunjukkan nilai *RMSE* 0.80 dan *R-squared* 0.99 (99%) dengan ini model regresi yang diuji memiliki kinerja yang sangat baik dalam memprediksi data.Hasil pengujian dapat dilihat pada Gambar 12.



Gambar 12. Visualisasi Hasil Pengujian

Berdasarkan hasil prediksi diatas, dapat dilihat bahwa sebagian besar titik-titik data terletak dekat dengan garis *"Perfect Fit"*. Ini menunjukkan bahwa model Regresi Linear cukup akurat dalam memprediksi suhu di Tokyo. Namun, ada beberapa titik yang sedikit jauh dari garis, yang mungkin menunjukkan adanya kesalahan prediksi. Secara keseluruhan, hasil analisis prediksi cuaca di Tokyo menggunakan algoritma Regresi Linear terlihat cukup baik. Grafik menunjukkan bahwa model memiliki tingkat akurasi yang mendekati sempurna.

Tabel 1. Evaluasi Model Prediksi

Algoritma	*RMSE*	*R-Squared*
Regresi Linear	0.80	0.99

Jika dilihat pada Tabel 1, model yang dihasilkan dari algoritma Regresi Linear memiliki nilai RMSE yang lebih rendah dan *R-Squared* yang lebih tinggi, menunjukkan bahwa model ini memiliki tingkat akurasi yang sangat baik dalam memprediksi cuaca di Tokyo untuk analisa perubahan suhu. Maka dapat dikatakan bahwa algoritma Regresi Linear mampu memberikan prediksi yang akurat khususnya dalam menganalisa perubahan suhu, karena nilai *R-Squared* yang mendekati 1 menunjukkan bahwa model dapat menjelaskan hampir seluruh variabilitas dalam data.

3.4 Pembahasan Hasil Penelitian

Pada penelitian ini menggunakan model prediksi Algoritma Regresi Linear dengan bahasa pemrograman *python* dan *tools Google Colab*. Setelah mendapatkan dataset, kemudian dilakukan tahapan *preprocessing* data dengan melakukan normalisasi data. Setelah data dinormalisasi, data diuji dengan metrik *RMSE (Root Mean Squared Error)* dan *R-squared*. RMSE mengukur seberapa dekat prediksi model dengan nilai sebenarnya, semakin kecil

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 2, February 2025, Page 553-562
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i2.6950

nilai RMSE semakin baik. *R-squared* mengukur seberapa baik model menjelaskan variasi data, semakin dekat nilai *R-squared* ke 1 semakin baik. Dari pengujian yang dilakukan menghasilkan nilai RMSE 0.80 dan nilai *R-Squared* 0.99. Maka dapat disimpulkan Algoritma Regresi Linear untuk memprediksi cuaca di Tokyo mendapatkan hasil akurasi yang cukup tinggi. Maka dari itu metode ini merupakan pilihan terbaik untuk prediksi cuaca.

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan oleh penulis, dimana penelitian ini bertujuan mengembangkan model prediksi cuaca di Tokyo menggunakan algoritma regresi linear, dengan fokus pada memprediksi suhu maksimum harian. Penelitian dilakukan menggunakan dataset historis cuaca selama 639 hari, menganalisis variabel seperti suhu, suhu minimum, kelembaban, curah hujan, dan tekanan udara. Model yang dikembangkan menunjukkan performa luar biasa dengan *Root Mean Squared Error (RMSE)* sebesar 0.80 dan koefisien determinasi *(R-squared)* mencapai 0.99, mampu menangkap hampir seluruh pola variabilitas cuaca. Hasil penelitian tidak hanya menunjukkan efektivitas regresi linear dalam prediksi cuaca, tetapi juga membuka peluang integrasi model ke dalam sistem prediksi cuaca yang lebih besar dan kompleks. Penelitian ini memberikan kontribusi signifikan terhadap pengembangan metode prediksi cuaca modern yang lebih akurat dan relevan dengan kebutuhan masyarakat perkotaan. Dengan prediksi cuaca yang lebih baik, masyarakat dapat mengambil langkah-langkah antisipasi yang tepat untuk menghadapi berbagai kondisi cuaca, seperti gelombang panas atau hujan lebat.

REFERENCES

[1] Y. Ohashi, T. Ihara, K. Oka, Y. Takane, and Y. Kikegawa, "Machine learning analysis and risk prediction of weather-sensitive mortality related to cardiovascular disease during summer in Tokyo, Japan," *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-44181-9.

[2] S. Ogata *et al.*, "Heatstroke predictions by machine learning, weather information, and an all-population registry for 12-hour heatstroke alerts," *Nat Commun*, vol. 12, no. 1, Dec. 2021, doi: 10.1038/s41467-021-24823-0.

[3] A. Marbun and D. Nofriansyah, "Analisa Data Mining Untuk Mengestimasi Potensi Curah Hujan Dengan Menggunakan Metode Regresi Linear Berganda," *Jurnal CyberTech*, vol. 4, no. 2, 2021

[4] M. Abdul, R. Wahid, A. Nugroho, and A. H. Anshor, "Prediksi Penyakit Kanker Paru-Paru Dengan Algoritma Regresi Linier," *Bulletin of Information Technology (BIT)*, vol. 4, no. 1, pp. 63–74, 2023, doi: 10.47065/bit.v3i1.

[5] A. Luthfiarta, A. Febriyanto, H. Lestiawan, and W. Wicaksono, "Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda," *JOINS (Journal of Information System)*, vol. 5, no. 1, pp. 10–17, May 2020, doi: 10.33633/joins.v5i1.2760.

[6] A. Supriyadi Sunge and A. Turmudi Zy, "Analisis Prediksi Penjualan Dengan Metode Regresi Linear Di Pt. Eagle Industry Indonesia," Jinteks, Vol 5, No 3, 2023, Available: https://journal.ekkeagle.com/jinteks/vol5/issue3/9910

[7] D. Rizki Septiani, "Pengembangan Model Prediksi Cuaca Menggunakan Teknik Machine Learning," *Jatika*, vol. 1, no. 4, pp. 1–16, 2024, doi: 10.33365/jatika.v4i1.2457.

[8] M. Edi, E. Utami, and A. Yaqin, "Prediksi Harga pada Trading Forex Pair USDCHF Menggunakan Regresi Linear," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 13, no. 2, pp. 109–119, Sep. 2023, doi: 10.34010/jamika.v13i2.9826.

[9] R. Andrianto and F. Irawan, "Implementasi Metode Regresi Linear Berganda Pada Sistem Prediksi Jumlah Tonase Kelapa Sawit di PT. Paluta Inti Sawit," *Jurnal Pendidikan Tambusai*, vol. 7–1, pp. 2926–2936, 2023, Accessed: Feb. 23, 2025

[10] J. P. Halawa, A. Hermawan, and . J., "Implementation of Linear Regression Algorithm to Predict Stock Prices Based on Historical Data," *bit-Tech*, vol. 5, no. 2, pp. 103–112, Dec. 2022, doi: 10.32877/bt.v5i2.616.

[11] A. Bahtiar, "Prediksi Hasil Panen Padi Tahun 2023 Menggunakan Metode Regresi Linier Di Kabupaten Indramayu," *Jurnal Informatika Terpadu*, vol. 9, no. 1, pp. 18–23, 2023, Available: https://journal.nurulfikri.ac.id/index.php/JIT

[12] Y. F. Wijaya and A. Triayudi, "Penerapan Data Mining Pada Prediksi Harga Emas dengan Menggunakan Algoritma Regresi Linear Berganda dan ARIMA," *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 1, pp. 73–81, Nov. 2023, doi: 10.47065/josyc.v5i1.4615.

[13] J. Pebralia, "Analisis Curah Hujan Menggunakan Machine Learning Metode Regresi Linier Berganda Berbasis Python dan Jupyter Notebook," JIFP (Jurnal Ilmu Fisika dan Pembelajarannya), vol. 6, no. 2, pp. 23–30, 2022, Available: http://jurnal.radenfatah.ac.id/index.php/jifp/

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 2, February 2025, Page 553-562
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i2.6950

[14] M. R. Alifi, H. Hayati, and C. Fauzi, "Penerapan Algoritma Regresi Linier pada Prediksi Tarif Influencer Media Sosial," *Journal of Information System Research (JOSH)*, vol. 4, no. 1, pp. 210–218, Oct. 2022, doi: 10.47065/josh.v4i1.2361.

[15] R. Ritonga, *Optimalisasi Kinerja Pegawai Pertanian*. PT. Literasi Nusantara Abadi Grup, 2024. Available: www.penerbitlitnus.co.id

[16] A. N. Sa'adah, A. S. Sunge, A. T. Zy, "Prediksi Pertumbuhan Penduduk Dengan Model Clustering Metode Regresi Linear," Jurnal Teknologi Terpadu, Vol 11, No 2, 2023, Available: https://journal.umm.ac.id/index.php/jtt/article/view/596894594

[17] T. Conradt, "Choosing multiple linear regressions for weather-based crop yield prediction with ABSOLUT v1.2 applied to the districts of Germany," *Int J Biometeorol*, vol. 66, no. 11, pp. 2287–2300, Nov. 2022, doi: 10.1007/s00484-022-02356-5.

[18] N. Afrilia and F. Frazna Az-Zahra, "Prediksi Hasil Panen Wortel Menggunakan Algoritma Regresi Linear Berganda," JATI, Vol 8, No 5, 2024, Available: https://ejournal.itn.ac.id/index.php/jati/article/view/10954

[19] T. A. Gahwera, O. Steven Eyobu, M. Isaac, and O. S. Eyobu, "Analysis of Machine Learning Algorithms for Prediction of Short-term Rainfall Amounts Using Uganda's Lake Victoria Basin Weather Dataset," *IEEE TRANSACTIONS and JOURNALS*, vol. vol 4, 2016, pp. 1–20, 2024, doi: 10.1109/ACCESS.2017.DOI.

[20] Z. Hadiansyah, Z. Rozikin, and M. Fatchan, "Implementasi Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Kanker Paru Paru," *Journal of Computer System and Informatics (JoSYC)*, vol. Vol 6, No. 1, pp. 96–106, Nov. 2024, doi: 10.47065/josyc.v6i1.6195.