

Implementasi *Data mining* Untuk Prediksi Panen Teh Menggunakan Metode *K-Nearest Neighbor* (KNN)

¹Muhammad Fahmi Hidayat, ²Bramantiyo Eko Putro

¹Teknik Industri, Universitas Suryakancana, Cianjur

²Teknik Industri, Universitas Suryakancana, Cianjur

E-mail: ¹hidayatfahmiaja@gmail.com, ²bramantiyoep91@gmail.com

ABSTRAK

Ketidakpastian hasil panen teh akibat variasi kondisi iklim menjadi tantangan dalam perencanaan produksi pada industri perkebunan teh. Perubahan curah hujan, suhu udara, kelembapan, dan intensitas penyinaran matahari dapat menyebabkan fluktuasi hasil panen. Penelitian ini bertujuan untuk memprediksi hasil panen teh berdasarkan variabel iklim menggunakan pendekatan *data mining* dengan algoritma *K-Nearest Neighbor* (KNN). Data yang digunakan berupa data sekunder yang terdiri dari data historis hasil panen teh dan data iklim, yang diproses melalui tahapan *pre-processing* berupa normalisasi data serta pembagian data latih dan data uji dengan rasio 80:20. Model prediksi dibangun menggunakan *KNN Regression* dan dievaluasi menggunakan *Mean Absolute Percentage Error* (MAPE) dan koefisien determinasi (R^2). Hasil evaluasi menunjukkan bahwa model KNN dengan parameter terbaik $k = 3$ menghasilkan nilai MAPE sebesar 39,67% dan nilai R^2 sebesar 0,000, yang menunjukkan bahwa akurasi prediksi numerik model masih tergolong rendah. Meskipun demikian, hasil prediksi menunjukkan adanya keterkaitan antara variabel iklim dan hasil panen, di mana kondisi iklim ekstrem cenderung diikuti oleh penurunan hasil panen teh. Model ini diharapkan dapat digunakan sebagai alat bantu awal dalam memahami kecenderungan hasil panen teh berbasis kondisi iklim.

Kata kunci : *Data mining, K-Nearest Neighbor, Prediksi Hasil Panen, Variabel Iklim, Teh.*

ABSTRACT

Uncertainty in tea yield caused by climatic variability poses a challenge for production planning in the tea plantation industry. Changes in rainfall, air temperature, humidity, and solar radiation intensity may lead to yield fluctuations. This study aims to predict tea yield based on climate variables using a data mining approach with the *K-Nearest Neighbor* (KNN) algorithm. The study uses secondary data consisting of historical tea yield data and climate data, which were processed through *pre-processing* stages including data normalization and an 80:20 split between training and testing data. The prediction model was developed using *KNN Regression* and evaluated using *Mean Absolute Percentage Error* (MAPE) and the coefficient of determination (R^2). The evaluation results indicate that the KNN model with the optimal parameter $k = 3$ produced a MAPE value of 39.67% and an R^2 value of 0.000, indicating relatively low numerical prediction accuracy. Nevertheless, the prediction results reveal a relationship between climate variables and tea yield, where extreme climatic conditions tend to be followed by a decrease in tea yield. This model is expected to serve as an initial analytical tool for understanding climate-based tea yield trends.

Keyword : *Data mining, K-Nearest Neighbor, Crop Yield Prediction, Climate Variables, Tea*

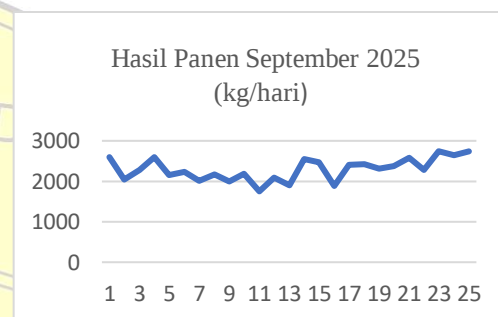
1. PENDAHULUAN

Industri teh merupakan salah satu sektor strategis yang berperan penting dalam perekonomian Indonesia sekaligus menjadi bagian dari warisan budaya bangsa. Namun, dalam beberapa tahun terakhir, sektor ini menghadapi tantangan serius terkait ketidakpastian hasil panen. Menurut Dinas Komunikasi dan Informatika Provinsi Jawa Barat (2025), produksi teh di Jawa Barat mengalami penurunan dari sekitar 62.000 ton pada tahun 2020 menjadi sekitar 55.000 ton pada tahun 2024 atau turun hampir 11%. Di Kabupaten Cianjur, produktivitas teh juga menurun dari rata-rata 1.200 kg pucuk basah per hektar menjadi sekitar 950–1.000 kg per hektar pada tahun 2024 akibat perubahan pola cuaca dan curah hujan yang tidak menentu.

Perubahan pola produksi teh di Kabupaten Cianjur mencerminkan adanya dinamika yang kompleks dalam sistem pertanian daerah tersebut. Fluktuasi ini menunjukkan bahwa sektor perkebunan teh sangat sensitif terhadap perubahan kondisi lingkungan dan manajemen produksi yang diterapkan. Ketidakpastian tersebut dipengaruhi oleh faktor eksternal seperti iklim, curah hujan, pola cuaca, serta teknik budidaya. Fluktuasi hasil panen tersebut berdampak langsung terhadap kualitas dan kuantitas daun teh yang tersedia, yang pada gilirannya memengaruhi stabilitas produksi dan daya saing produk di pasar global (Chlingaryan dkk., 2018).

Teknologi *data mining* menawarkan pendekatan yang efektif untuk menganalisis pola historis data produksi, iklim, maupun faktor lingkungan sehingga prediksi hasil panen dapat dilakukan secara lebih akurat (Zhu & Ghoshray, 2021). Salah satu metode yang banyak digunakan adalah *K-Nearest Neighbor* (KNN), yang mampu mengklasifikasikan data berdasarkan kedekatan karakteristik dengan data sebelumnya. Implementasi metode KNN di bidang pertanian juga telah menunjukkan hasil yang baik dalam memprediksi hasil panen, misalnya pada komoditas padi, yang membuktikan bahwa metode ini cukup andal digunakan untuk mengantisipasi ketidakpastian produksi.

PT. Tenggara Perkebunan Teh Maleber, sebagai salah satu produsen teh terkemuka yang beroperasi di wilayah Pacet, Cianjur, turut merasakan dampak serius. PT. Tenggara mengelola total area perkebunan seluas sekitar 100 hektar dan berfokus pada produksi teh Hitam dan teh *Oolong*. Ketergantungan perusahaan terhadap pasokan daun teh segar yang konsisten dari kebunnya sangat tinggi. Akan tetapi ketidakpastian hasil panen membuat pemenuhan kebutuhan pasokan daun teh segar menjadi sulit. Fluktuasi hasil panen dapat dilihat pada Gambar 1.



Gambar 1. Hasil panen teh September 2025

Pada Gambar 1. menunjukkan adanya fluktuasi produksi yang cukup signifikan selama 25 hari pengamatan, dengan hasil terendah sebesar 1.751 kg/hari dan tertinggi mencapai 2.743 kg/hari. Secara umum, produksi pada awal periode cenderung menurun hingga pertengahan waktu, kemudian meningkat kembali pada hari-hari terakhir. Pola ini mengindikasikan bahwa hasil panen teh dipengaruhi oleh faktor-faktor eksternal seperti kondisi cuaca, kesuburan tanah, serta siklus pertumbuhan pucuk teh yang tidak seragam. Meskipun demikian, rata-rata hasil panen berada pada kisaran 2.300–2.400 kg/hari, yang mencerminkan tingkat produktivitas yang relatif stabil dan masih tergolong baik untuk skala perkebunan teh dengan variabilitas alamiah.

Masalah ketidakpastian hasil panen yang dialami PT. Tenggara Perkebunan Teh Maleber menunjukkan perlunya upaya peningkatan ketepatan dalam memperkirakan hasil produksi. Oleh karena itu, diperlukan suatu penelitian yang memanfaatkan metode *data mining*, khususnya *K-Nearest Neighbor* (KNN), untuk membantu perusahaan dalam memprediksi hasil panen secara lebih akurat.

Dengan adanya prediksi yang lebih tepat, perusahaan dapat mengambil keputusan produksi dan perencanaan operasional secara lebih efektif.

2. LANDASAN TEORI

2.1 Tanaman Teh (*Camellia sinensis*)

Tanaman teh (*Camellia sinensis*) merupakan komoditas perkebunan yang memiliki nilai ekonomi tinggi dan banyak dibudidayakan di wilayah dataran tinggi Indonesia. Pertumbuhan tanaman teh dipengaruhi oleh kondisi lingkungan seperti curah hujan, suhu udara, kelembapan, dan penyinaran matahari. Faktor-faktor tersebut berperan dalam menentukan produktivitas tanaman teh dan hasil panen yang dihasilkan setiap periode produksi (Anjarsari dkk., 2020).

Selain digunakan sebagai bahan minuman, teh juga memiliki manfaat kesehatan yang luas karena kandungan polifenol, katekin, dan antioksidannya berperan dalam menjaga daya tahan tubuh dan mencegah penyakit degeneratif (Mathew & Negi, 2021).

2.2 Faktor Iklim terhadap Produktivitas Teh

Produktivitas tanaman teh sangat dipengaruhi oleh kondisi iklim yang secara langsung memengaruhi proses fisiologis tanaman, khususnya pembentukan dan pertumbuhan pucuk (Siahaan dkk., 2024). Curah hujan yang stabil membantu proses pertumbuhan daun teh, sementara suhu udara memengaruhi aktivitas fotosintesis tanaman. Kelembapan udara berperan dalam menjaga keseimbangan fisiologis tanaman, sedangkan lama penyinaran matahari menentukan proses pembentukan energi melalui fotosintesis. Oleh karena itu, perubahan kondisi iklim dapat menyebabkan fluktuasi hasil panen teh dari waktu ke waktu.

2.3 Data mining

Data mining merupakan proses menemukan pola atau pengetahuan baru dari kumpulan data dalam jumlah besar menggunakan teknik statistik, matematika, dan *machine learning* (Han, 2012). Dalam bidang pertanian dan perkebunan, *data mining* dapat digunakan untuk melakukan analisis

prediktif, termasuk memprediksi hasil produksi berdasarkan data historis dan variabel lingkungan.

2.3.1 Pre-Processing

Sebagian besar data mentah yang tersimpan dalam database umumnya belum dapat langsung digunakan karena masih bersifat tidak lengkap, tidak terstruktur, dan mengandung *noise*. Misalnya, terdapat data yang sudah tidak relevan, duplikasi, memiliki nilai yang hilang (*missing value*), mengandung pencilan (*Outlier*), atau memiliki format yang tidak sesuai dengan kebutuhan proses *data mining*.

Selain itu, sering ditemukan pula nilai data yang tidak konsisten dengan ketentuan atau kebijakan yang berlaku. Oleh karena itu, sebelum digunakan dalam proses analisis, data perlu melalui tahap *data preprocessing*, yang meliputi pembersihan data (*data cleaning*) dan transformasi data, agar data yang diolah menjadi lebih konsisten, akurat, serta siap digunakan untuk proses *data mining* (Larose & Larose, 2014).

2.4 Algoritma K-Nearest Neighbor (KNN) Regression

K-Nearest Neighbor (KNN) adalah metode *supervised learning* non-parametrik yang memprediksi nilai target untuk data baru berdasarkan rata-rata (atau rata-rata tertimbang) dari *k* tetangga terdekat dalam ruang fitur (Sitienei dkk., 2023). Salah satu ukuran jarak yang umum digunakan adalah *Euclidean Distance*. Metode KNN memiliki kelebihan dalam kemampuannya mengenali pola dari data historis tanpa memerlukan pembentukan model matematis yang kompleks.

3. METODOLOGI

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *data mining* untuk memprediksi hasil panen teh berdasarkan variabel iklim. Model prediksi dibangun menggunakan algoritma *K-Nearest Neighbor* (KNN) Regression, yang bekerja dengan menentukan nilai prediksi berdasarkan kedekatan data baru terhadap data historis dalam ruang fitur. Tahapan metodologi

penelitian dilakukan secara sistematis sebagai berikut:

3.1 Pengumpulan Data

Data yang digunakan merupakan data sekunder dari PT Tenggara Perkebunan Teh Maleber yang terdiri dari data iklim dan data hasil panen teh bulanan. Variabel yang digunakan dalam penelitian meliputi suhu minimum, suhu maksimum, curah hujan, kecepatan angin, lama penyinaran matahari, kelembapan udara, dan hasil panen teh sebagai variabel target.

3.2 Pre-processing Data

Pre-processing Data dilakukan untuk memastikan kualitas data sebelum proses pemodelan. Tahap ini meliputi pemeriksaan kelengkapan data, deteksi nilai ekstrem (*outlier*), serta normalisasi data menggunakan metode *Min-Max Normalization* untuk menyamakan skala antarvariabel.

3.3 Pemodelan KNN Regression

Pada tahap ini dilakukan proses pembentukan model prediksi menggunakan algoritma KNN Regression. Nilai prediksi ditentukan berdasarkan kedekatan jarak antar data menggunakan *Euclidean Distance*. Parameter jumlah tetangga terdekat (nilai k) ditentukan melalui pengujian beberapa nilai untuk memperoleh model dengan performa terbaik.

3.4 Evaluasi Model

Evaluasi model dilakukan untuk menilai tingkat keakuratan dan keandalan hasil prediksi yang dihasilkan oleh algoritma KNN dibandingkan dengan data aktual (Hodson, 2022). Pengukuran performa model menggunakan *Mean Absolute Percentage Error* (MAPE) dan koefisien determinasi (R^2) untuk mengetahui tingkat kesalahan prediksi dan kemampuan model dalam menjelaskan variasi data.

4. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini terdiri dari data hasil panen teh serta data faktor lingkungan yang diduga berpengaruh terhadap variasi produksi. Data hasil panen diperoleh dari catatan operasional harian PT Tenggara Perkebunan Teh Maleber, di mana

jumlah pucuk teh yang dipetik dicatat setiap hari kerja. Pada perkebunan ini, kegiatan panen tidak dilakukan pada hari libur, sehingga hanya hari kerja yang memiliki data panen.

4.1 Pre-Processing

1. Handling Missing Value

Berdasarkan hasil pemeriksaan, tidak ditemukan *missing value* pada seluruh variabel yang digunakan dalam penelitian ini. Seluruh data historis tercatat secara lengkap dan konsisten pada setiap periode observasi. Dengan demikian, tidak diperlukan proses imputasi data, penghapusan baris (data deletion), maupun teknik penanganan lainnya yang umum digunakan dalam pengolahan data yang tidak lengkap.

2. Identify Outlier

Row No.	outlier	Suhu Min	Suhu Maks	Curah Hujan...	KecepatanA...	SinarMatah...	Kelembapan...	Hasil Panen
1	false	18	25	420	2	4	88	19873
2	false	18	25	380	2	4	87	19106
3	false	18	25	310	2	5	85	20637
4	false	19	26	240	2	5	83	17581
5	false	19	26	180	2	6	80	19873
6	false	18	27	110	2	7	76	18956
7	false	18	27	90	3	7	74	17982
8	false	18	27	85	3	7	73	20637
9	false	18	27	120	3	7	75	21004
10	false	18	27	210	2	6	80	18786
11	false	19	26	340	2	5	85	22016
12	false	18	25	410	2	4	88	22990
13	false	18	25	424	2	4	88	26584
14	false	18	25	383	2	4	87	24038
15	false	18	25	313	2	5	85	27606

Gambar 2. Identify Outlier

Berdasarkan hasil tersebut, dari seluruh data yang dianalisis terdapat 10 data yang teridentifikasi sebagai *Outlier* dan 107 data lainnya termasuk dalam kategori data normal. Keberadaan *Outlier* terutama dipengaruhi oleh variasi nilai pada variabel hasil panen dan curah hujan yang memiliki rentang nilai cukup besar.

3. Normalisasi Data

Hasil normalisasi data menggunakan metode *Z-Transformation* menghasilkan nilai atribut dalam bentuk skor baku, sebagaimana ditunjukkan pada Tabel IV.4. Nilai setelah normalisasi memiliki rentang negatif dan positif dengan rata-rata mendekati nol, yang menunjukkan bahwa seluruh variabel telah berhasil ditransformasikan ke dalam skala yang sebanding dan siap digunakan pada tahap pemodelan *K-Nearest Neighbor*.

Row No.	Suhu Min	Suhu Maks	Curah Hujan...	Kecelakaan ...	Sinar Matah...	Kelambapan...	Hasil Panen
1	-0.932	-1.466	1.280	-0.578	-1.327	1.272	-1.284
2	-0.932	-1.466	0.961	-0.578	-1.327	1.087	-1.384
3	-0.932	-1.466	0.457	-0.578	-0.492	0.716	-1.184
4	1.063	-0.399	-0.067	-0.578	-0.492	0.345	-1.587
5	1.063	-0.399	-0.516	-0.578	0.344	-0.211	-1.284
6	-0.932	0.968	-1.040	-0.578	1.179	-0.953	-1.404
7	-0.932	0.968	-1.190	1.713	1.179	-1.324	-1.532
8	-0.932	0.968	-1.227	1.713	1.179	-1.510	-1.184
9	-0.932	0.968	-0.965	1.713	1.179	-1.139	-1.136
10	-0.932	0.968	-0.291	-0.578	0.344	-0.211	-1.429
11	1.063	-0.399	0.682	-0.578	-0.492	0.716	-1.003
12	-0.932	-1.466	1.206	-0.578	-1.327	1.272	-0.915
13	-0.932	-1.466	1.310	-0.578	-1.327	1.272	-0.406
14	-0.932	-1.466	1.004	-0.578	-1.327	1.087	-0.673
15	-0.932	-1.466	0.480	-0.578	-0.492	0.716	-0.271

Gambar 3. Normalisasi Data

4.2 Pemodelan K-Nearest Neighbor (KNN) Regression



Gambar 4. Model KNN Regression

Berdasarkan hasil perhitungan *Mean Absolute Percentage Error* (MAPE) pada berbagai variasi nilai k , diperoleh MAPE sebesar 81,98% untuk $k = 1$, 39,67% untuk $k = 3$, dan 50,85% untuk $k = 5$. Nilai MAPE yang tinggi menunjukkan bahwa terdapat perbedaan yang cukup besar antara nilai prediksi model dengan nilai aktual hasil panen. Secara umum, semakin kecil nilai MAPE maka semakin baik kemampuan model dalam melakukan prediksi. Oleh karena itu, model dengan $k = 3$ dipilih sebagai model terbaik karena menghasilkan nilai MAPE paling rendah dibandingkan dengan variasi k lainnya.

4.3 Hasil Prediksi

Row No.	Suhu Min	Suhu Maks	Curah Hujan...	Kecelakaan ...	Sinar Matah...	Kelambapan...	prediction...	Hasil Panen
1	18.000	26.000	128.000	3	7.000	75.000	25655.133	49515.000
2	18.000	27.000	224.000	2	6	80.000	25476.250	40978.000
3	18.000	26.000	363.000	2	5.000	85.000	25004.950	49515.000
4	18.000	26.000	453.000	2	4	88.000	26236.608	40889.000
5	18.000	25.000	410.000	2	4	87.000	37245.384	38958.000
6	18.000	26.000	259.000	2	5.000	83.000	25452.405	41967.000
7	18.000	27.000	194.000	2	6	80.000	26036.370	43221.000
8	18.000	27.000	118.000	2	7.000	76.000	24610.633	32243.000
9	18.000	27.000	87.000	3	7.000	74.000	25651.844	41916.000
10	18.000	26.000	91.000	3	7.000	73.000	25481.833	33648.000
11	18.000	26.000	129.000	3	7.000	75.000	25655.104	33169.000
12	18.000	27.000	226.000	2	6	80.000	25517.133	26367.000
13	18.000	26.000	367.000	2	5.000	85.000	25056.500	37903.000
14	18.000	26.000	442.000	2	4	88.000	26958.801	38741.000
15	18.000	26.000	457.000	2	4	88.000	26349.163	41531.000
16	18.000	25.000	414.000	2	4	87.000	37334.816	43208.000

Gambar 5. Hasil Prediksi KNN Regression

Berdasarkan hasil pemodelan *K-Nearest Neighbor* (KNN) Regression dengan

pembagian data sebesar 80% sebagai data pelatihan dan 20% sebagai data pengujian menggunakan metode *linear sampling*, dari Gambar 5. diperoleh nilai prediksi hasil panen yang secara umum berada pada rentang 25.000-37.000 kg per periode, sedangkan nilai hasil panen aktual berada pada kisaran yang lebih luas, yaitu sekitar 28.000-49.000 kg. Penggunaan *linear sampling* memastikan bahwa data pengujian merupakan data pada periode waktu terakhir, sehingga hasil prediksi mencerminkan kemampuan model dalam mengestimasi hasil panen pada kondisi yang benar-benar belum pernah dilihat sebelumnya.

4.4 Evaluasi Model

Evaluasi performa model *K-Nearest Neighbor* (KNN) Regression dilakukan menggunakan data uji untuk mengukur akurasi prediksi hasil panen teh secara objektif. Penilaian kinerja model menggunakan beberapa metrik, yaitu MAE, MSE, RMSE, dan koefisien determinasi (R^2), yang digunakan untuk mengukur tingkat kesalahan prediksi serta kemampuan model dalam menjelaskan hubungan antara nilai aktual dan nilai prediksi.

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 1.246 +/- 0.000
absolute_error: 0.983 +/- 0.766
squared_error: 1.552 +/- 1.969
squared_correlation: 0.000
```

Gambar 6. Evaluasi Model

1. Mean Absolute Error (MAE)

Hasil evaluasi model menggunakan RapidMiner menunjukkan nilai *Mean Absolute Error* (MAE) sebesar 0,983. Nilai MAE tersebut merepresentasikan rata-rata selisih absolut antara nilai hasil panen aktual dan nilai prediksi yang dihasilkan oleh model *K-Nearest Neighbor* (KNN) Regression. Nilai MAE yang relatif kecil menunjukkan bahwa model memiliki tingkat kesalahan prediksi yang rendah pada data yang telah dinormalisasi, sehingga prediksi yang dihasilkan cenderung mendekati nilai aktual.

2. Mean Squared Error (MSE)

Nilai *Mean Squared Error* (MSE) yang diperoleh sebesar 1,552. Metrik ini menunjukkan rata-rata kuadrat selisih antara nilai aktual dan nilai prediksi, sehingga kesalahan dengan deviasi besar akan memberikan pengaruh yang lebih signifikan. Nilai MSE tersebut mengindikasikan bahwa masih terdapat beberapa kesalahan prediksi yang cukup tinggi pada kondisi data tertentu, meskipun secara umum model mampu mengikuti pola data hasil panen.

3. *Root Mean Squared Error* (RMSE)

Hasil evaluasi menunjukkan bahwa *Root Mean Squared Error* (RMSE) sebesar 1,246. RMSE merupakan akar kuadrat dari MSE yang merepresentasikan rata-rata kesalahan prediksi dalam satuan yang sama dengan variabel target. Nilai RMSE yang lebih tinggi dibandingkan MAE menunjukkan adanya beberapa deviasi prediksi yang relatif besar, sehingga metrik ini memberikan gambaran yang lebih konservatif terhadap tingkat kesalahan model secara keseluruhan.

4. *Coefficient of Determination* (R^2)

Hasil evaluasi menunjukkan nilai *Coefficient of Determination* (R^2) sebesar 0,000, yang mengindikasikan bahwa model KNN *Regression* belum mampu menjelaskan variasi data hasil panen berdasarkan variabel prediktor yang digunakan. Nilai tersebut menunjukkan bahwa hubungan antara nilai prediksi dan nilai aktual masih sangat lemah, sehingga variabel lingkungan yang digunakan belum cukup merepresentasikan pola hasil panen secara memadai.

5. KESIMPULAN

Penelitian ini bertujuan untuk membangun model prediksi hasil panen teh menggunakan metode *K-Nearest Neighbor* (KNN) *Regression* berdasarkan variabel lingkungan yang meliputi curah hujan, suhu minimum, suhu maksimum, kelembapan udara, lama penyinaran matahari, dan kecepatan angin. Proses *pre-processing* menunjukkan bahwa dataset yang digunakan tidak mengandung *missing value*, sehingga model dibangun menggunakan data yang lengkap dan konsisten.

Hasil evaluasi menunjukkan bahwa model terbaik diperoleh pada parameter $k = 3$ dengan

nilai MAPE sebesar 39,67%, yang mengindikasikan tingkat kesalahan prediksi masih relatif tinggi. Nilai evaluasi lainnya menunjukkan MAE sebesar 0,983, MSE sebesar 1,552, dan RMSE sebesar 1,246, yang mencerminkan adanya deviasi prediksi pada beberapa periode pengamatan. Selain itu, nilai *Coefficient of Determination* (R^2) sebesar 0,000 menunjukkan bahwa model belum mampu menjelaskan variasi hasil panen secara signifikan berdasarkan variabel lingkungan yang digunakan.

Secara keseluruhan, penelitian ini menunjukkan bahwa variabel iklim memiliki hubungan dengan fluktuasi hasil panen teh, namun belum cukup kuat untuk menghasilkan model prediksi dengan tingkat akurasi yang tinggi. Oleh karena itu, diperlukan penambahan variabel lain seperti faktor agronomis, kondisi tanah, manajemen pemeliharaan, atau variabel produksi internal perusahaan untuk meningkatkan performa model prediksi pada penelitian selanjutnya.

6. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada PT Tenggara Perkebunan Teh Maleber yang telah memberikan izin penelitian serta menyediakan data yang diperlukan dalam penelitian ini. Ucapan terima kasih juga disampaikan kepada dosen pembimbing yang telah memberikan arahan, masukan, dan bimbingan selama proses penelitian hingga penyusunan artikel ini. Selain itu, penulis mengapresiasi semua pihak yang telah membantu secara langsung maupun tidak langsung dalam pelaksanaan penelitian ini..

DAFTAR PUSTAKA

- Anjarsari, I. R. D., Ariyanti, M., & Rosniawaty, S. (2020). Studi ekofisiologis tanaman teh guna meningkatkan pertumbuhan, hasil, dan kualitas teh. *Kultivasi*, 19(3). <https://doi.org/10.24198/kultivasi.v19i3.26623>
- Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches

- for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, 151, 61–69.
<https://doi.org/10.1016/j.compag.2018.05.012>
- Data Mining. (2012). Elsevier.
<https://doi.org/10.1016/C2009-0-61819-5>
- Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, 15(14), 5481–5487.
<https://doi.org/10.5194/gmd-15-5481-2022>
- Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining* (1st ed.). Wiley.
<https://doi.org/10.1002/9781118874059>
- Mathew, N. S., & Negi, P. S. (2021). Phenolic content and anti-oxidative attributes of various parts of wild banana (*Ensete superbum* Roxb. Cheesman) plant. *Journal of Food Biochemistry*, 45(4).
<https://doi.org/10.1111/jfbc.13657>
- Siahaan, A. J., Irsal, & Sitorus, B. (2024). The Relationship between Climate and Productivity of Tea Plants (*Camellia sinensis* L.) in Bah-Butong Gardens PT. Perkebunan Nusantara IV, North Sumatra Province in 2005-2009. *Jurnal Online Pertanian Tropik*, 11(2), 1–8.
<https://doi.org/10.32734/jopt.v11i2.17559>
- Sitienei, M., Otieno, A., & Anapapa, A. (2023). An Application of K-Nearest-Neighbor Regression in Maize Yield Prediction. *Asian Journal of Probability and Statistics*, 24(4), 1–10.
<https://doi.org/10.9734/ajpas/2023/v24i4529>
- Zhu, Y., & Ghoshray, A. (2021). *Climate Anomalies and Their Impact on Cereal Grain Prices*.
<https://doi.org/10.22004/AG.ECON.315271>