

Prediction of Palm Oil Production Results PT.KSI Solok Selatan Using *Ensemble k-Nearest Neighbor (kNN)*

Nilda Yanti, Atus Amadi Putra*, Dony Permana, dan Zilrahmi

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: atusamadiputra@fmipa.unp.ac.id

Submitted : 14 Desember 2023

Revised : 31 Januari 2024

Accepted : 16 Februari 2024

ABSTRACT

PT. KSI experienced production decrease that the cause of replanting that happened in 2022. In managing palm oil production PT. KSI has problems with palm oil production results not reaching out the targets so it can affect the Company's Work Plan and Budget, therefore it is very necessary to predict palm oil production results so that all palm oil production and processing activities can run according to plan. The ensemble technique is a method that is capable of making accurate predictions and is used very effectively in the kNN method, therefore there is no need to search for the best k value. Based on the results of the analysis that has been carried out, it can be seen that by using an ensemble the level of accuracy is 9.36%, which is considered high accuracy compared to just using a single kNN with k = 1 of 10.84%. So it can be concluded that the model has worked well with the data.

Keywords: *Ensemble, k-NN, Oil Palm, Production, Prediction.*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Kelapa sawit adalah tanaman yang menjadi komoditas utama perkebunan di Indonesia. Daerah Solok Selatan memiliki beberapa perusahaan kelapa sawit baik dari segi kepemilikan pemerintah ataupun swasta. Berdasarkan data statistik perkebunan Indonesia tentang kelapa sawit dari Direktorat Jenderal Perkebunan (2022), produksi kelapa sawit Indonesia pada tahun 2020 sebesar 45,741 juta ton dan naik menjadi 48,235 juta ton di tahun 2022. Sedangkan produksi komoditas tanaman lain, seperti karet 3,135 juta ton, kelapa 2,859 juta ton dan kopi sebesar 793 ribu ton.

PT. Kencana Sawit Indonesia (KSI) adalah sebuah perusahaan milik swasta yang berkecimpung dalam bagian perkebunan produksi kelapa sawit dan industri pengolahan kelapa sawit. Pada tahun 2022 PT. KSI mengalami penyusutan produksi kelapa sawit karena dilakukan langkah replanting. Dalam pengelolaan produksi kelapa sawit PT. KSI ini memiliki kendala pada hasil produksi kelapa sawit yang tidak mencapai target, sehingga berdampak pada Rencana Kerja dan Anggaran Perusahaan. Oleh karena itu, untuk mengantisipasi penurunan produksi kelapa sawit perlu dilakukan prediksi hasil produksi kelapa sawit.

Terdapat berbagai macam metode yang dapat digunakan untuk prediksi data *time series* diantaranya adalah *k-Nearest Neighbor (kNN)*. Menurut kang (2021) metode *kNN* tidak memerlukan asumsi khusus mengenai distribusi data yang harus dipenuhi, sehingga metode ini lebih fleksibel dan efektif untuk diimplementasikan. Metode ini bekerja dengan mencari titik data terdekat (disebut sebagai "*neighbors*") dari sebuah titik data baru yang ingin diprediksi. Kemudian, nilai dari titik data baru akan ditentukan berdasarkan mayoritas nilai tetangga terdekatnya. Metode *kNN* awalnya merupakan suatu metode yang digunakan untuk menganalisis klasifikasi, namun belakangan ini juga digunakan untuk prediksi (Alfin, 2022). Dalam beberapa dekade metode *kNN* ini populer digunakan untuk prediksi.

Algoritma *kNN* adalah salah satu metode yang banyak digunakan dalam penghitungan data *mining*, dengan cara kerja yang sederhana namun memiliki tingkat akurasi yang tinggi, serta juga andal terhadap data *training* sampel yang *noisy* (Virdaus, 2020). Metode *kNN* ini juga mudah untuk di implementasikan. Walaupun metode *kNN* ini sederhana dan mudah, namun metode *kNN* mempunyai keunggulan dibandingkan dengan metode lainnya, yaitu kumpulan data *training* yang relatif kecil dapat digeneralisasi (Satriya dkk, 2018). Hasil prediksi dipengaruhi oleh pemilihan nilai *k*, penentuan nilai *k* sangat penting dalam metode *kNN*. Menurut Sinta (2015) nilai *k* yang kecil dapat menghasilkan *varian* yang besar, sedangkan nilai *k* besar dapat menyebabkan bias model yang besar. Alternatif metode yang biasa digunakan

untuk menangani permasalahan tersebut adalah dengan mengoptimasi parameter dengan cara *cross-validation*, namun metode optimasi dengan *cross-validation* ini kurang efisien karena algoritma pelatihan harus diulang pada k berikutnya. Oleh sebab itu diperlukan optimasi dengan membuat atau melengkapi suatu metode yang dapat menghasilkan hitungan yang lebih baik. Salah satu cara yang dapat digunakan adalah dioptimasi dengan menggunakan metode *Ensemble* pada k NN.

Salah satu penelitian terdahulu yang dilakukan oleh Jusman (2022) dengan judul metode ensemble k -nearest neighbor untuk prediksi indeks harga saham gabungan (IHSG). Dari penelitian tersebut diperoleh nilai MAPE sebesar 7.16%, yang mana ini menunjukkan bahwa penggunaan metode *ensemble k*NN dalam memprediksi memberikan hasil yang cukup akurat.

Dari penjelasan tersebut, maka dilakukanlah penelitian yang bertujuan untuk mengetahui akurasi metode *k*-Nearest Neighbor yang dioptimasi dengan Teknik *ensemble* dalam memprediksi hasil produksi kelapa sawit di PT.KSI Solok Selatan.

II. METODE PENELITIAN

A. Jenis Penelitian dan Sumber Data

Jenis penelitian yang digunakan dalam penelitian ini adalah penelitian terapan yang digunakan untuk memprediksi hasil produksi kelapa sawit di PT.KSI Solok Selatan dengan menggunakan metode *Ensemble k*-Nearest Neighbor (k NN). Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh langsung dari staf bagian perkebunan PT.KSI Solok Selatan. Penelitian ini menggunakan data hasil produksi kelapa sawit di PT.KSI dari periode Januari 2018 – Desember 2022.

B. Teknik Analisis Data

Untuk menganalisis data dalam penelitian ini digunakan dengan bantuan *software* Rstudio. Berikut tahapan analisis data yang dilakukan

1. Mengumpulkan data langsung ke PT.KSI Solok Selatan
2. Membagi dataset menjadi data *training* dan data *testing*

Pada penelitian ini dataset dibagi menjadi data *training* dan data *testing*. Adapun rasio pembagian data *training* dan data *testing* yang digunakan yaitu 80% : 20% (Rogel, 2018). Data dari bulan Januari 2018 hingga Desember 2021 sebagai data *training* dan data dari bulan Januari sampai Desember 2022 sebagai data *testing*.

3. Mengkonversi data dari univariat menjadi multivariat

Dalam metode k NN, setiap data terdiri dari variabel fitur dan variabel target. Menurut Martinez dkk (2017) untuk menerapkan k NN pada pengaturan prediksi *time series* data univariat terlebih dahulu harus dikonversi ke bentuk data multivariat agar bisa digunakan. Mengkonversi data dari univariat menjadi multivariat merupakan mengubah data yang awalnya hanya memiliki satu variabel menjadi data yang memiliki dua variabel atau lebih. Konversi data ini dapat dilakukan dengan menambahkan fitur baru pada dataset. Fitur baru ini dibuat berdasarkan nilai waktu sebelumnya atau yang biasa disebut dengan lag (nilai kelambanan) dan nilai kelambanan ini dapat disusun menjadi kolom-kolom baru.

Dalam data *time series* jika panjang periode musiman adalah m , maka akan digunakan lag $1:m$. Contohnya untuk data bulanan maka digunakan lag $1:12$. Secara sederhana konversi data dapat digambarkan sebagai berikut. Misalkan terdapat data *time series* bulanan selama 5 tahun, yaitu $t = \{x_1, x_2, x_3, x_4, x_5, \dots, x_{60}\}$. Maka digunakan lag $1:12$ variabel fitur. Berikut hasil konversi data diberikan

Tabel 1. Stuktur Data Setelah Dikonversi

No	Fitur	Target
1	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}$	x_{13}
2	$x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}$	x_{14}
⋮	⋮	⋮
3	$x_{48}, x_{49}, x_{50}, x_{51}, x_{52}, x_{53}, x_{54}, x_{55}, x_{56}, x_{57}, x_{58}, x_{59}$	x_{60}

4. Melakukan tranformasi data

Dalam *time series* transformasi data dilakukan untuk meningkatkan kinerja model dan akurasi peramalan. Salah satu transformasi yang dapat digunakan adalah transformasi *additive* (Martinez dkk., 2022) menggunakan persamaan

$$f_x^{i'} = f_x^i - \frac{1}{n} \sum_{i=1}^n f_x^i \quad (1)$$

Keterangan:

f_x^i : Nilai fitur ke- i data ke- x

$f_x^{i'}$: Hasil transformasi nilai fitur ke- i data ke- x

n : Banyaknya fitur

$$t_x' = t_x - \frac{1}{n} \sum_{i=1}^n f_x^i \quad (2)$$

Keterangan:

t_x : Nilai target data ke- x

t_x' : Hasil transformasi nilai target data ke- x

f_x^i : Nilai fitur ke- i data ke- x

n : Banyaknya fitur

5. Memilih beberapa nilai k untuk model k NN

Pada penelitian terdapat 36 nilai k yang tersedia dan akan digunakan sebagai model, yaitu $k = 1$ sampai $k = 36$. Alasan menggunakan seluruh nilai k karena ingin melihat performa seluruh nilai k .

6. Menghitung jarak *Euclidean*

Salah satu yang populer digunakan yaitu jarak *Euclidean* yang dapat dihitung menggunakan persamaan berikut

$$d(f^i, q) = \sqrt{\sum_{x=1}^n (f_x^i - q_x)^2} \quad (3)$$

Keterangan:

f_x^i : nilai fitur ke- x baris ke- i dari data *training*

q_x : nilai fitur ke- x baris ke- i dari data yang akan diprediksi

$d(f^i, q)$: jarak *Euclidean* antara data *training* ke- i dengan data yang akan diprediksi

n : banyak data *training*

7. Menghitung nilai prediksi model k NN

Data *time series* merupakan data yang memperhatikan urutan waktu, maka data *training* yang terpilih sebagai k tetangga terdekat dapat diberi bobot. Tujuannya adalah untuk memberikan bobot lebih kepada tetangga yang lebih dekat. Pemberian bobot menggunakan Persamaan

$$\hat{y} = \frac{\sum_{i=1}^k w_i g_i}{\sum_{i=1}^k w_i} \quad (4)$$

Keterangan:

\hat{y} : Nilai prediksi

w_i : Nilai pembobot untuk tetangga terdekat ke- i

g_i : Nilai target tetangga terdekat ke- i

k : banyaknya tetangga terdekat

Untuk memboboti nilai target yang terpilih menjadi tetangga terdekat digunakan nilai pembobot. Untuk menghitung nilai pembobot digunakan Persamaan

$$w_i = \frac{1}{d_i^2} \quad (5)$$

Keterangan:

w_i : Nilai pembobot untuk tetangga terdekat ke- i yang terpilih

d_i : Jarak *Euclidean* antara data yang akan diprediksi dengan tetangga ke- i

8. Mentransformasi hasil prediksi ke bentuk awal

Untuk mentransformasi hasil prediksi ke bentuk awal, rata-rata vektor fitur dari data yang akan diprediksi ditambahkan kedalam hasil prediksi. Secara matematis dapat dilihat pada Persamaan

$$\hat{y}' = \hat{y} + \frac{1}{n} \sum_{i=1}^n q_x \tag{6}$$

Keterangan:

- \hat{y} : Hasil prediksi (dalam bentuk transformasi *additive*)
- \hat{y}' : Hasil prediksi setelah ditransformasi kebentuk awal
- n : Banyaknya fitur
- q_x : Nilai fitur ke- x dari data yang akan diprediksi

9. Menggabungkan hasil prediksi menggunakan teknik *ensemble*

Saat melakukan prediksi, performa metode k NN tunggal dapat dioptimalkan dengan menggunakan metode *ensemble*. Teknik *ensemble* ini sangat penting untuk meningkatkan kemampuan akurasi prediksi dari berbagai model standar. Teknik *ensemble* merupakan teknik yang tidak memilih satu model dari beberapa kandidat model yang kemudian melakukan estimasi dari model tersebut, tetapi teknik *ensemble* ini menyatukan hasil estimasi dari beberapa model yang ada dengan bobot tertentu.

Ada beberapa macam jenis teknik *ensemble* yang dapat digunakan salah satunya adalah *simple averaging*, di mana teknik ini menyatukan hasil prediksi dari masing-masing model dengan cara menjumlahkannya dan kemudian membaginya dengan jumlah model, seperti yang dijelaskan dalam Persamaan

$$\mu = \frac{1}{m} \sum_{i=1}^m \hat{y}_i \tag{7}$$

Keterangan:

- \hat{y}_i : Nilai prediksi dari model ke- i
- μ : Hasil *ensemble*
- m : Banyaknya model k NN yang digunakan

10. Evaluasi menggunakan MAPE

Pendekatan MAPE ini berguna untuk mengevaluasi ketepatan dalam prediksi. Semakin kecil nilai MAPE, maka hasil prediksi akan semakin akurat.

$$MAPE = \frac{1}{n} \sum \left| \frac{X_t - F_t}{F_t} \right| \times 100\% \tag{8}$$

Keterangan:

- X_t : Nilai aktual ke- t
- F_t : Nilai prediksi dari model ke- t
- n : Banyaknya fitur

Berikut tabel akurasi hasil prediksi yang diperoleh berdasarkan nilai MAPE (Maricar, 2019)

Tabel 2. Akurasi Nilai MAPE

Nilai MAPE	Akurasi
$MAPE \leq 10\%$	Tinggi
$10\% < MAPE < 20\%$	Baik
$20\% < MAPE < 50\%$	Sedang
$MAPE > 50\%$	Rendah

III. HASIL DAN PEMBAHASAN

Hasil produksi kelapa sawit PT.KSI Solok selatan yang di prediksi dengan menggunakan metode *ensemble k-nearest neighbor* ini dianalisis dengan bantuan *software* Rstudio. Langkah pertama dalam analisis ini adalah *dataset* dibagi menjadi data *training* dan data *testing*, yaitu data dari bulan Januari 2018 sampai Desember 2021 sebagai data *training* dan bulan Januari sampai Desember 2022 sebagai data *testing*. Setelah *dataset* tadi dibagi langkah selanjutnya yaitu mengubah data *training* menjadi bentuk multivariat yang dapat dilihat pada tabel 3.

Tabel 3. Stuktur Data *Training* Setelah Dikonversi

No	Fitur	Target
1	13.522, 12.229, 18.038, 16.879, 20.438, 17.941, 18.902, 15.877, 15.174, 19.615, 18.792, 17.135	23.969
2	12.229, 18.038, 16.879, 20.438, 17.941, 18.902, 15.877, 15.174, 19.615, 18.792, 17.135, 23.969	14.565
⋮	⋮	⋮
36	16.373, 15.214, 13.616, 12.832, 12.248, 14.513, 13.103, 15.480, 15.391, 15.826, 17.420, 15.012	14.299

Berdasarkan Tabel 3. Dapat dilihat bahwa struktur data *training* setelah dikonversi menghasilkan 36 baris data yang terdiri dari variabel fitur dan juga variabel target. Setelah konversi data maka akan dilakukan tranformasi yang dapat dilihat pada tabel 4.

Tabel 4. Hasil Transformasi Data *Training*

No	Fitur	Target
1	-4023.16, -5316.16, 492.83, -666.16, 2892.83, 395.83, 1356.83, 4331.83, -2371.16, 2069.83, 1246.83, -410.16	6423.83
2	-6186.75, -377.75, -1536.75, 2022.25, -474.75, 486.25, 3461.25, -3241.75, 1199.25, 376.25, -1280.75, 5553.25	-3850.75
⋮	⋮	⋮
36	1615.08, 5246.08, 331.08, 21.08, 13924.08, -1317.91, -5040.91, -966.91, -3066.91, -6614.91, -2421.91, -1707.91	-9064.91

Berdasarkan Tabel 4 terlihat bahwa hasil transformasi data *training* menghasilkan nilai positif dan negatif. Transformasi ini dilakukan untuk meningkatkan akurasi peramalan. Langkah selanjutnya yaitu menghitung jarak *Euclidean* antara data *training* dengan data yang akan diprediksi. Berikut hasil jarak *Euclidean*.

Tabel 5. Jarak *Euclidean*

No	1	2	3	4	...	37
1	0.000	8296.210	9984.541	7628.963	...	8034.090
2	8296.210	0.000	13856.336	10767.597	...	10035.336
3	9984.541	13856.336	0.000	3590.736	...	11602.286
4	7628.963	10767.597	3590.736	0.000	...	8451.850
⋮	⋮	⋮	⋮	⋮	⋮	⋮
37	17937.086	18701.625	13583.784	12457.862	...	0.000

Berdasarkan Tabel 5 dapat dilihat bahwa jarak *Euclidean* setiap baris data. Elemen di baris pertama dan kolom kedua menunjukkan jarak *Euclidean* antara baris pertama dan kedua dari data asli. Namun, karena jarak *Euclidean* simetris (yaitu jarak *Euclidean* antara baris 1 dan 2 sama dengan jarak antara baris 2 dan 1), maka elemen di kolom pertama dan baris kedua juga menunjukkan jarak yang sama. Jarak *Euclidean* antara baris dengan dirinya sendiri selalu nol, sehingga elemen diagonal utama (yaitu elemen di baris pertama dan kolom pertama, baris kedua dan kolom kedua, dan seterusnya) selalu nol.

Langkah selanjutnya memilih beberapa nilai k untuk model k NN. Dalam penelitian ini digunakan nilai k yang terdiri dari $k=1$, $k=11$, $k=12$, dan $k=13$. Kemudian dilakukan prediksi untuk periode ke depan. Berikut hasil prediksinya

Tabel 6. Hasil Prediksi

No	Bulan	$k=1$	$k=11$	$k=12$	$k=13$	<i>Ensemble</i>
1	Januari	24.396	19.253	19.231	19.355	20.559
2	Februari	15.017	18.511	18.197	18.015	17.435
3	Maret	14.428	19.524	20.231	19.911	18.524
4	April	13.039	19.184	20.279	20.107	18.152
5	Mei	16.810	20.521	21.454	21.173	19.990
6	Juni	19.283	21.097	21.134	20.716	20.557
7	Juli	21.990	20.383	21.677	20.856	21.226

No	Bulan	$k=1$	$k=11$	$k=12$	$k=13$	Ensemble
8	Agustus	17.396	22.049	22.676	22.307	21.107
9	September	18.863	22.136	22.394	22.164	21.389
10	Oktober	14.708	23.032	23.639	23.145	21.131
11	November	14.111	22.990	23.543	23.397	21.010
12	Desember	15.009	23.583	23.181	22.923	21.174

Pada Tabel 6 terlihat bahwa nilai-nilai hasil prediksi dengan menggunakan beberapa nilai k , yaitu $k=1$, $k=11$, $k=12$, dan $k=13$ serta yang dioptimasi dengan *ensemble*. Pada nilai prediksi yang di optimasi dengan *ensemble* terlihat bahwa hasil produksi kelapa sawit mengalami penurunan pada bulan february yaitu diperkirakan sekitar 17.435 kg lalu cenderung naik hingga ke bulan Desember.

Untuk melihat keakuratan hasil prediksi dapat dilihat menggunakan nilai MAPE. MAPE digunakan untuk melihat sejauh mana prediksi model mendekati nilai sebenarnya pada data testing. Berikut merupakan tabel nilai MAPE.

Tabel 7. Nilai MAPE

k NN	MAPE (%)	Akurasi
1	10,84	Baik
11	10,87	Baik
12	11,10	Baik
13	11,24	Baik
Ensemble	9,36	Tinggi

Berdasarkan tabel 7, dapat terlihat bahwa performa model *ensemble* memiliki nilai MAPE terkecil diantara semua model yaitu sebesar 9,36%. Sedangkan untuk model k NN tunggal hanya memberikan nilai MAPE terkecil sebesar 10,84% dengan nilai $k=1$. Penggunaan *ensemble* pada k NN menunjukkan bahwa akurasi dapat meningkat.

IV. KESIMPULAN

Pada penelitian dilakukan prediksi hasil produksi kelapa sawit PT.KSI Solok Selatan dengan menggunakan metode *Ensemble k-Nearest Neighbor*. Berdasarkan hasil analisis yang telah dilakukan, dapat terlihat bahwa dengan menggunakan *ensemble* tingkat akurasi sebesar 9,36% yang termasuk kedalam akurasi yang tinggi dibandingkan dengan hanya menggunakan k NN tunggal dengan $k=1$ sebesar 10,84%. Sehingga hal ini dapat disimpulkan bahwa model telah bekerja dengan baik terhadap data karena nilai MAPE < 10%. Saran untuk penelitian selanjutnya karena dalam penelitian menggunakan teknik *ensemble* menggunakan *simple averaging*, jadi penelitian selanjutnya dapat menggunakan teknik *ensemble* lain.

DAFTAR PUSTAKA

- Alfin, N. F. (2022). Penerapan Metode Clustering untuk Prediksi Produksi Bawang Merah (Ensemble K-Nearest Neighbor). *Vol 3*, 30-37.
- Jamil, A. (n.d.). *Statistik Perkebunan Unggulan Nasional 2020-2022*. Jakarta: Sekretariat Direktorat Jenderal Perkebunan. Retrieved from <https://ditjenbun.pertanian.go.id/template/uploads/2022/08/STATISTIK-UNGGULAN-2020-2022.pdf>
- Jusman, M., Nur'eni, N., & Handayani, L. (2022). Metode Ensemble K-Nearest Neighbor untuk Prediksi Indeks Harga Saham Gabungan (IHSG) di Indonesia. *Jurnal Matematika, Statistika dan Komputasi*, 18(3), 423–433.
- Kang, S. (2021). k-Nearest Neighbor Learning with Graph Neural Networks. *Mathematics*, 9(8)
- Maricar, M. A. (2019, Mei). Analisis Perbandingan Nilai Akurasi Moving Average dan Exponential Smoothing untuk Sistem Peramalan Pendapatan Pada Perusahaan XYZ. *Jurnal Sistem dan Informatika*, Vol. 13, No. 2, 36-45.
- Martínez, F., Frías, M. P., Pérez, M. D., & Rivera, A. J. (2017). A Methodology For Applying k-Nearest Neighbor To Time Series Forecasting. *Artificial Intelligence Review*, 52(3), 2019–2037. <https://link.springer.com/article/10.1007/s10462-017-9593-z>

- Martínez, F., Charte, F., Frías, M. P., & Martínez-Rodríguez, A. M. (2022). Strategies for Time Series Forecasting with Generalized Regression Neural Networks. *Neurocomputing*, 491, 509–521. <https://doi.org/10.1016/j.neucom.2021.12.028>
- Nanja, M. (2015, Februari). Metode K-Nearest Neighbor Berbasis Forward Selection untuk Prediksi Harga Komoditi Lada. *Jurnal Pseudocode, Volume 2*.
- Rogel, S. J. (2018). *Data Science and Analytics with Python*. Chapman and Hall/CRC.
- Satriya, S. E. (2018). Implementasi Metode Ensemble K-Nearest Neighbor untuk Prediksi Nilai Tukar Rupiah Terhadap Dollar Amerika. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol. 2 No. 4*, 1718-1725.
- Sinta, D. (2015). *Metode Ensemble K-Nearest Neighbor Untuk Prediksi Harga Beras Di Indonesia*. Bogor: Institut Pertanian Bogor.
- Virdaus, D., & Prasetyaningrum, P. T. (2020). Penerapan Data Mining Untuk Memprediksi Harga Bawang Merah Di Yogyakarta Menggunakan Metode K-Nearest Neighbor. *Journal Of Information System And Artificial Intelligence, 1(1)*, 1-8.