



IMPLEMENTASI WEIGHTED K-NEAREST NEIGHBOR UNTUK PERAMALAN DATA DERET WAKTU

AZKIA NADWAH, NUSYIRWAN, AHMAD FAISOL*

Jurusan Matematika, FMIPA Universitas Lampung, Bandar Lampung, Indonesia

*email : ahmadfaisol@fmipa.unila.ac.id

ABSTRAK

Peramalan merupakan suatu situasi atau kondisi yang diperkirakan akan terjadi pada masa yang akan datang. Dalam meramalkan data deret waktu dibutuhkan metode peramalan yang efisien untuk menghasilkan nilai peramalan yang akurat. Pada penelitian ini Metode Weighted k-Nearest Neighbor (WK-NN) dipilih dengan tujuan untuk menerapkan metode WK-NN untuk peramalan data deret waktu dan meninjau kinerja yang dihasilkan. Data yang digunakan pada penelitian ini adalah data Produk Domestik Bruto (PDB) kehutanan dan penebangan kayu dengan kurun waktu Triwulan I 2010 hingga Triwulan IV 2019. Hasil penelitian menunjukkan bahwa metode WK-NN memberikan kinerja terbaik pada data PDB kehutanan dan penebangan kayu dengan $k = 13$ dan nilai Mean Absolute Percentage Error (MAPE) yang diperoleh adalah 0.0038%.

Kata Kunci: Data Deret Waktu, Metode Weighted k-Nearest Neighbor (WK-NN), Mean Absolute Percentage Error (MAPE).

ABSTRACT

Forecasting is a situation or condition that is expected to occur in the future. In forecasting time series data, an efficient forecasting method is needed to produce accurate forecasting values. In this study, the Weighted k-Nearest Neighbor (WK-NN) method was chosen to apply the WK-NN method for forecasting time series data and review the resulting performance. The data used in this study is forestry and logging Gross Domestic Product (GDP) data from the first quarter of 2010 to the fourth quarter of 2019. The results show that the WK-NN method provides the best performance on forestry and logging GDP data with $k = 13$, and the Mean Absolute Percentage Error (MAPE) value obtained is 0.0038%.

Keywords: Time Series Data, Weighted k-Nearest Neighbor (WK-NN) Method, Mean Absolute Percentage Error (MAPE).

1 Pendahuluan

Peramalan merupakan suatu situasi atau kondisi yang diperkirakan akan terjadi pada masa yang akan datang. Peramalan pada dasarnya menggunakan data masa lampau yang

dianalisis menggunakan cara-cara tertentu. Data masa lampau dikumpulkan, dipelajari dan dianalisis dihubungkan dengan perubahan waktu [1]. Dalam meramalkan data deret waktu dibutuhkan metode peramalan yang efisien untuk menghasilkan nilai peramalan yang akurat. Banyak metode yang berkembang untuk meramalkan data deret waktu, salah satunya adalah metode K-Nearest Neighbor (K-NN). K-Nearest Neighbor merupakan salah satu metode untuk mengambil keputusan menggunakan pembelajaran terawasi dimana hasil dari data masukan yang baru diklasifikasi berdasarkan terdekat dalam data nilai [2].

Metode K-NN merupakan metode untuk mencari sejumlah k objek yang memiliki jarak terdekat antara data latih dengan data uji. Data latih merupakan subset data yang digunakan untuk melatih algoritma dibentuk sesuai lag yang diterapkan, sedangkan data uji merupakan data yang digunakan untuk menguji algoritma yang sudah dilatih. Sangat umum untuk memperkirakan lebih dari satu nilai di masa mendatang. Untuk tujuan ini, strategi Multiple Input Multiple Output (MIMO) dipilih [3]. Pembentukan data latih menggunakan strategi MIMO untuk membentuk data latih input sebagai data latih yang akan dihitung jaraknya dengan data uji dan data latih target sebagai data latih yang digunakan untuk peramalan.

Kelemahan pada metode K-NN adalah tidak adanya pengaruh jarak pada perumusan peramalan serta pada metode K-NN biasa, bobot setiap data pada k data terdekat besarnya sama yaitu $\frac{1}{k}$ [4]. Pemberian bobot yaitu invers kuadrat jarak pada metode K-NN dapat memperbaiki tingkat kesalahan [5]. Pemberian bobot pada K-NN disebut juga Weighted K-Nearest Neighbor (WK-NN). WK-NN pada dasarnya memiliki prinsip yang sama dengan metode K-NN. Pada metode WK-NN invers kuadrat dari jarak Euclid digunakan sebagai bobot, kemudian bobot tersebut digabungkan dengan algoritma metode K-NN.

Oleh karena itu, peneliti termotivasi untuk meramalkan data deret waktu menggunakan metode WK-NN untuk mendapatkan hasil ramalan yang lebih sistematis dan akurat. Penelitian ini bertujuan untuk meninjau kinerja metode WK-NN dalam meramalkan data deret waktu. Indikator keakuratan peramalan dilihat sesuai nilai galat yang akan dihitung menggunakan Mean Absolute Percentage Error (MAPE). Semakin kecil nilai galat yang didapatkan maka hasil peramalan semakin baik.

2 Metode

Sumber data pada penelitian ini diambil dari *website* Badan Pusat Statistika (BPS). Data tersebut berupa data Produk Domestik Bruto (PDB) kehutanan dan penebangan kayu dari Triwulan I 2010 hingga Triwulan IV 2019. Penelitian ini akan melakukan peramalan data deret waktu yaitu PDB kehutanan dan penebangan kayu pada periode Triwulan I 2020 hingga Triwulan IV 2020. Metode yang diusulkan untuk peramalan tersebut adalah WK-NN. Langkah-langkah peramalan menggunakan WK-NN adalah sebagai berikut:

a) Pembentukan Data

Data dibagi menjadi dua bagian yaitu data latih dengan data uji. Pembentukan data latih menggunakan strategi *Multiple Input Multiple Output* (MIMO), pola data terakhir pada data PDB kehutanan dan penebangan kayu dijadikan data uji. Strategi MIMO merupakan strategi untuk membentuk target untuk horizon perkiraan dalam simulasi tunggal dengan model unik. MIMO mempertimbangkan satu set data sebagai berikut:

$$M = \{(x_i^{latih}, y_i^{latih}) \in (R^d \times R^H)\}_{i=1}^N$$

dengan

$$\begin{aligned} x_i^{latih} &\subset \{x_i, x_{i+1}, \dots, x_{i+d-1}\} \\ y_i^{latih} &= \{x_{i+d}, x_{i+d+1}, \dots, x_{i+d+H-1}\} \end{aligned}$$

Strategi MIMO membatasi semua horizon untuk diprediksi (H) dengan struktur model yang sama, misalnya dengan set input (d = jumlah lag) yang sama. Hal ini akan sangat menghambat fleksibilitas pendekatan prediksi dan secara berlebihan dapat membiaskan model yang dikembalikan. Ini bukan kasus dari strategi Langsung, di mana setiap model diizinkan untuk menggunakan set variabel input yang berbeda [3].

b) Mengurutkan Jarak

Salah satu ukuran kedekatan yang digunakan adalah jarak Euclid antara dua titik yaitu titik pada data latih dan titik pada uji, dirumuskan sebagai berikut [6]:

$$D_{i,j} = \sqrt{\sum_{i=1}^n (x_i^{uji} - x_i^{latih})^2}$$

dengan:

- $D_{i,j}$ = jarak Euclid
- x^{latih} = data latih input
- x^{uji} = data uji
- n = banyaknya data

Jarak Euclid diurutkan berdasarkan jarak terdekat hingga jarak terjauh dari data latih input ke data uji. Untuk dapat melihat pola data ke berapa yang memiliki kemiripan paling mendekati dengan data uji.

c) Implementasi *Weighted K-Nearest Neighbor* (WK-NN)

Metode WK-NN merupakan pengembangan dari metode K-NN. Metode WK-NN memiliki prinsip yang sama dengan metode K-NN yaitu mencari jarak terdekat antara data uji dengan sejumlah k data terdekatnya dalam data latih. WK-NN akan memberi bobot terberat pada data terdekatnya dan terkecil pada data terjauh sesuai fungsi jarak [7].

$$w_{i,j} = \frac{1}{D_{i,j}^2}$$

dengan invers jarak kuadrat sebagai pembobot peramalan WK-NN, sehingga terbentuklah perumusan peramalan sebagai berikut:

$$F_t = \frac{\sum_{i=1}^k w_{i,t} y_i^{latih}}{\sum_{i=1}^k w_{i,t}}$$

dengan:

- F_t = hasil peramalan ke t
- $w_{i,t}$ = bobot setiap data latih target
- y_i^{latih} = data latih target ke i

d) Evaluasi Hasil Peramalan

Salah satu indikator yang dapat digunakan untuk menghitung galat dalam peramalan adalah *Mean Absolute Percentage Error* (MAPE). Evaluasi hasil peramalan dihitung berdasarkan nilai MAPE yang diperoleh, semakin kecil nilai MAPE maka hasil peramalan semakin akurat. MAPE dapat dihitung dengan persamaan berikut [7]:

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{x_t - F_t}{x_t} \right|}{n} \times 100$$

dengan:

MAPE= nilai galat hasil peramalan

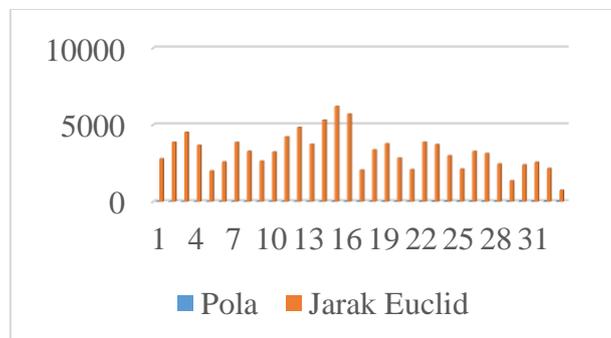
x_t = data aktual pada periode t

F_t = peramalan pada periode t

n= jumlah data ramalan

3 Hasil dan Pembahasan

Proses pembentukan data latih menggunakan strategi MIMO dengan $d = 4$ dan $H = 4$ menghasilkan 33 pola data latih input dan target. Data PDB kehutanan dan penebangan kayu Triwulan I 2019 hingga Triwulan IV 2019 menjadi data uji setiap data latih input yang terbentuk. Pembentukan data latih dengan menggunakan strategi MIMO pada data PDB kehutanan dan penebangan kayu menjadikan 4 triwulan berturut-turut sebagai data latih input dan 4 triwulan selanjutnya berturut-turut setelah data input sebagai data latih target begitu seterusnya hingga data terakhir pada data latih. Jarak antara data latih input dengan data uji disajikan pada Gambar 1 sebagai berikut:



Gambar 1. Jarak Euclid setiap pola data.

Berdasarkan Gambar 1, jarak terdekat diperoleh oleh data latih target Triwulan I hingga Triwulan IV tahun 2019 yaitu pola ke 33, pada urutan kedua jarak terdekat diperoleh oleh data latih target Triwulan I hingga Triwulan IV tahun 2018 yaitu pola ke 29, dan pada urutan ketiga diperoleh oleh data latih target Triwulan I hingga Triwulan IV tahun 2012 yaitu pola ke 5. Jarak terjauh diperoleh oleh data latih target Triwulan IV tahun 2014 yaitu pola ke 15 dan Triwulan I hingga Triwulan III tahun 2015 yaitu pola ke 16.

Penentuan nilai k yang digunakan sesuai pola data latih yang terbentuk yaitu 33 pola data latih. Sehingga, nilai k yang akan digunakan adalah $k = 1, k = 2, k = 3, k = 4, k = 5, k = 6, k = 7, k = 8, k = 9, k = 10$, dan seterusnya hingga $k = 33$. Melakukan peramalan dengan WK-NN dibutuhkan nilai bobot pada setiap data latih target yaitu menggunakan invers kuadrat dari jarak Euclid. Bobot tersebut dapat dihitung dengan persamaan $w_{i,j}$.

$$w_{1,1} = \frac{1}{D_{1,1}^2}$$

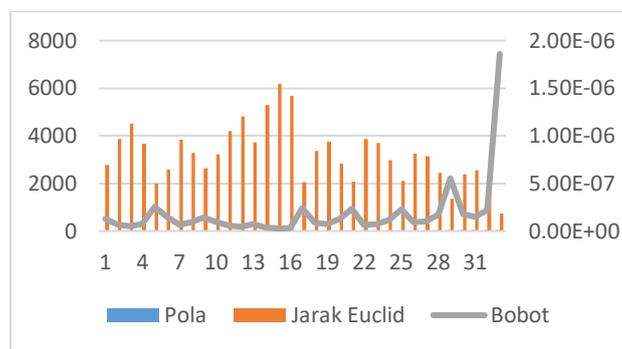
$$w_{2,1} = \frac{1}{D_{2,1}^2}$$

$$w_{3,1} = \frac{1}{D_{3,1}^2}$$

$$\vdots$$

$$w_{33,1} = \frac{1}{D_{33,1}^2}$$

Berdasarkan perhitungan diatas diperoleh nilai invers setiap jarak Euclid yang menghasilkan kesimpulan semakin jauh jarak antara data latih dengan data uji, akan semakin kecil data tersebut terboboti. Begitupun sebaliknya, semakin dekat jarak antara data latih dengan data uji, akan semakin besar data tersebut terboboti.



Gambar 2. Hasil perhitungan nilai bobot.

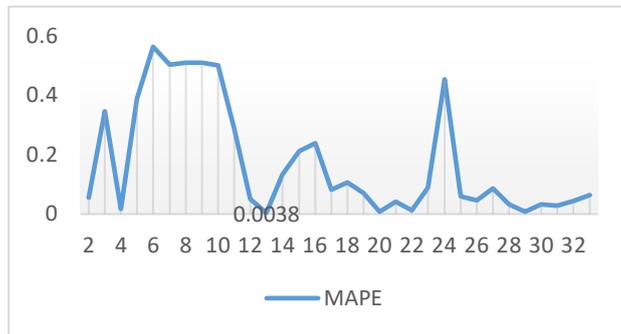
Berdasarkan Gambar 2, dapat dilihat pada pola data ke 33 nilai bobot sangat besar dan memiliki nilai jarak yang sangat dekat. Begitu pula dengan pola-pola data yang memiliki jarak Euclid yang kecil seperti pada pola data ke 5, 9, 17, 21, 25, dan 29. Pola-pola data tersebut memiliki nilai bobot yang dihasilkan lebih besar dari yang lainnya, dibandingkan dengan nilai bobot pada pola data ke 15 yang memiliki nilai jarak yang sangat tinggi.

Dalam penelitian ini nilai invers setiap jarak Euclid dibobotkan pada y_i^{latih} . Proses pembobotan untuk metode WK-NN menghasilkan model Peramalan data PDB kehutanan dan penebangan kayu Triwulan I 2020 hingga Triwulan IV disesuaikan berdasarkan nilai k yang ditentukan. Untuk model peramalan Triwulan I dengan nilai $k = 2$ dengan y_{33}^{latih} dan y_{29}^{latih} sesuai urutan jarak terdekat pertama dan kedua didapatkan model peramalan sebagai berikut:

$$F_1 = \frac{\sum_{i=1}^2 w_{i,t} y_i^{latih}}{\sum_{i=1}^2 w_{i,t}} = \frac{(w_{1,1} \times y_{33}^{latih}) + (w_{2,1} \times y_{29}^{latih})}{(w_{1,1} + w_{2,1})}$$

Begitupula dengan model peramalan Triwulan I dengan nilai $k = 3$ data latih yang digunakan adalah y_{33}^{latih} , y_{29}^{latih} dan y_5^{latih} sesuai urutan jarak terdekat pertama, kedua dan ketiga. Proses pemodelan peramalan tersebut terus berlanjut seperti pemodelan diatas hingga peramalan menggunakan nilai $k = 33$. Berlaku pula untuk peramalan Triwulan II, III, dan IV dengan $k = 2$ hingga $k = 33$.

Setiap hasil peramalan pada jumlah data terdekat menghasilkan nilai galat yang dihitung menggunakan MAPE. Pemilihan peramalan terbaik dilakukan menggunakan ukuran ketepatan peramalan yaitu MAPE yang disajikan pada Gambar 3 sebagai berikut:



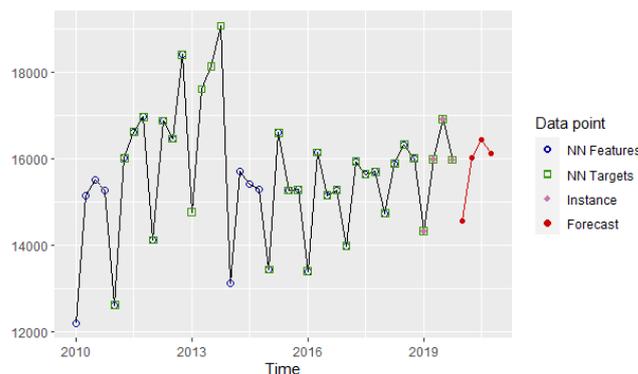
Gambar 3. Mean Absolute Percentage Error (MAPE)

Berdasarkan pada Gambar 3, yang merupakan gambar rangkuman nilai MAPE dari setiap peramalan dengan nilai k dimulai dari 2 sampai 33. Didapatkan peramalan terbaik dengan MAPE terkecil 0.0038% dengan nilai k pada WK-NN sama dengan 13. Kinerja metode WK-NN dilihat berdasarkan nilai MAPE, semakin kecil nilai MAPE yang diperoleh maka akan semakin baik kinerja WK-NN pada peramalan data PDB kehutanan dan penebangan kayu. Berikut adalah hasil peramalan terbaik PDB kehutanan dan penebangan kayu untuk Triwulan I hingga Triwulan IV tahun 2020.

Tabel 1. Peramalan PDB kehutanan dan penebangan kayu dengan $k=13$

Waktu	Peramalan
Triwulan I 2020	14569.14
Triwulan II 2020	16024.04
Triwulan III 2020	16453.34
Triwulan IV 2020	16115.06

Dari Tabel 1 dapat dilihat hasil peramalan selama satu tahun kedepan dengan PDB kehutanan dan penebangan kayu mengalami kenaikan pada Triwulan I hingga Triwulan II, kemudian mengalami penurunan pada Triwulan IV. Dari hasil peramalan tersebut juga dapat diketahui PDB kehutanan dan penebangan kayu paling sedikit pada Triwulan I dan paling banyak pada Triwulan III. Berikut ini grafik peramalan PDB kehutanan dan penebangan kayu dengan metode WK-NN.



Gambar 4. Peramalan PDB kehutanan dan penebangan kayu Triwulan I-Triwulan IV 2020

Pada Gambar 4, menggambarkan hasil proses peramalan menggunakan metode WK-NN dengan $k = 13$. Simbol lingkaran berwarna biru merupakan data latih input yang terpilih, simbol kotak berwarna hijau merupakan data latih target yang terpilih, dan simbol titik merah muda merupakan data uji. Garis yang berwarna hitam merupakan plot data PDB kehutanan dan penebangan kayu dari Triwulan I tahun 2010 sampai Triwulan IV 2019. Garis yang berwarna merah merupakan plot nilai peramalan PDB kehutanan dan penebangan kayu dari Triwulan I sampai Triwulan IV tahun 2020.

4 Kesimpulan

Kinerja yang dihasilkan dari penerapan WK-NN untuk menyelesaikan peramalan data deret waktu dengan mempertimbangkan pola kemiripan data dan dengan nilai k yang digunakan menghasilkan nilai peramalan dengan galat yang berbeda-beda pada masing-masing periode peramalan. Berdasarkan peramalan terbaik pada data PDB kehutanan dan penebangan kayu setelah melakukan penelitian, nilai k yang optimal untuk data tersebut adalah $k = 13$ dengan galat MAPE 0.0038%.

Daftar Pustaka

- [1] S. Makridakis, S.C. Wheelwriht, and McGee. "Metode dan Aplikasi Peramalan," Terjemahan Hari Suminto, Binarupa Aksara, Jakarta, 1999.
- [2] Teknomo, K. (2006). What is K-Nearest Neighbor Algorithm ? Dipetik 06 22, 2016, dari <http://people.revolude.com/kardi/tutorial/KNN/What-isK-Nearest-Neighbor-Algorithm.html>.
- [3] S.B. Taieb, G. Bontempi, A. Sorjamaa, and A. Lendasse, "Long-term prediction of time series by combining direct and MIMO strategies," Proceedings of the 2009 IEEE International Joint Conference on Neural Networks, USA, pp. 3054-3061, 2009.
- [4] S.B. Setiawan, Adiwijaya, and M.S. Mubarok, "Klasifikasi Topik Berita Berbahasa Indonesia Menggunakan Weighted K-Nearest Neighbor," *e-Proceeding of Engineering.*, vol. 5, no. 1, pp. 1819-1825, 2018.
- [5] D.T. Larose, "Discovering Knowledge in Data an Introduction to Data Mining". Jhon Wiley & Sons, Inc., New Jersey, 2005.
- [6] B. Santosa, "Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis," Graha Ilmu, Yogyakarta, 2007.
- [7] F. Martinez, M.P. Frias, F. Charte, and A.J. Rivera, "Time Series Forecasting with KNN in R:the tsfknn Package," The R Journal., vol. 11, no. 2, pp. 229-242, 2019.