

Bab VIII

Kompleksitas Pemodelan Deret Waktu untuk Cuaca: Kasus Kelembapan Relatif dan Penyinaran Matahari

Januardi

A. Urgensi Pemodelan Iklim/Iklim sebagai Sistem yang Kompleks

Cuaca dan iklim mempunyai pengaruh besar terhadap kehidupan di Bumi. Mereka adalah bagian dari pengalaman sehari-hari manusia dan penting untuk kesehatan, produksi pangan, dan kesejahteraan. Banyak pihak yang menganggap bahwa perubahan iklim yang disebabkan oleh manusia sebagai suatu hal yang memprihatinkan. Laporan penilaian kedua yang dilakukan IPCC menyajikan bukti ilmiah bahwa aktivitas manusia mungkin telah memengaruhi iklim. Jika seseorang ingin memahami, mendeteksi, dan pada akhirnya memprediksi pengaruh manusia terhadap iklim, kita perlu mema-

Januardi

Universitas Padjadjaran, *e-mail*: januardi@unpad.ac.id

© 2024 Editor & Penulis

Januardi. (2024). Kompleksitas pemodelan deret waktu untuk cuaca: Kasus kelembapan relatif dan penyinaran matahari. Dalam D. E. Nuryanto & I. Fathrio (Ed.), *Prediksi iklim untuk ketahanan pangan* (225–261). Penerbit BRIN. DOI: 10.55981/brin.1244.c1392. E-ISBN: 978-602-6303-49-3.

hami sistem yang menentukan iklim bumi dan proses-proses yang menyebabkan perubahan iklim.

Variasi dan perubahan iklim, yang disebabkan oleh faktor eksternal, mungkin dapat diprediksi terutama pada skala spasial yang lebih besar, yaitu benua dan global. Karena aktivitas manusia, seperti emisi gas rumah kaca atau perubahan penggunaan lahan, diyakini bahwa aspek skala besar dari perubahan iklim yang disebabkan oleh manusia juga dapat diprediksi. Namun, kemampuan untuk melakukan hal tersebut terbatas karena kita tidak dapat secara akurat memprediksi perubahan populasi, perubahan ekonomi, perkembangan teknologi, dan karakteristik relevan lainnya dari aktivitas manusia pada masa depan. Oleh karena itu, dalam praktiknya, kita harus bergantung pada skenario perilaku manusia yang disusun secara cermat dan menentukan proyeksi iklim berdasarkan skenario tersebut.

Banyak proses interaksi fisik, kimia, dan biologi yang terjadi di antara berbagai komponen sistem iklim dalam rentang skala ruang dan waktu yang luas sehingga menjadikan sistem tersebut menjadi sangat kompleks. Meskipun komponen-komponen sistem iklim sangat berbeda dalam komposisi, sifat fisik dan kimia, struktur dan perilakunya, semuanya saling terkait oleh fluks massa, panas, dan momentum (semua subsistem bersifat terbuka dan saling berhubungan). Banyak metode yang saat ini telah berkembang dan digunakan untuk memahami dan menganalisis iklim, seperti hierarki model sistem bumi satu dimensi yang sederhana hingga empat dimensi yang rumit atau memisahkan berbagai proses yang terlibat dalam iklim. Metode-metode tersebut terus berkembang mengingat adanya kebutuhan prediksi iklim dan cuaca untuk menunjang berbagai aktivitas ekonomi, terutama di sektor pertanian dan pangan.

Iklim memberikan pengaruh penting pada proses pertumbuhan dan pemanenan dari tanaman pangan. Tingginya suhu dan rendahnya kelembapan udara akan menyebabkan laju evaporasi yang berlebihan pada air di tanah (Heino et al., 2023). Akibatnya, nutrisi yang ada pada air akan ikut teruapkan dan tidak dapat diserap oleh akar tanaman. Laju pertumbuhan tanaman akan terhambat dan panen yang diha-

silkan lebih sedikit. Perubahan iklim dewasa ini yang bersuhu tinggi dengan kelembapan yang rendah membuat tingginya probabilitas suatu daerah menuju kekeringan (Tabari & Willems, 2023). Juga tingginya pola ketidakpastian (*uncertainty*) dari perubahan iklim (Jeantet et al., 2023) membuat tantangan dalam proses penanaman dan pemanenan tanaman pangan menjadi lebih sulit. Oleh karena itu, pengembangan teknik, teknologi, dan model prediksi cuaca dan iklim sangat diperlukan untuk proses penjadwalan penanaman dan pemanenan tanaman pangan yang lebih optimal.

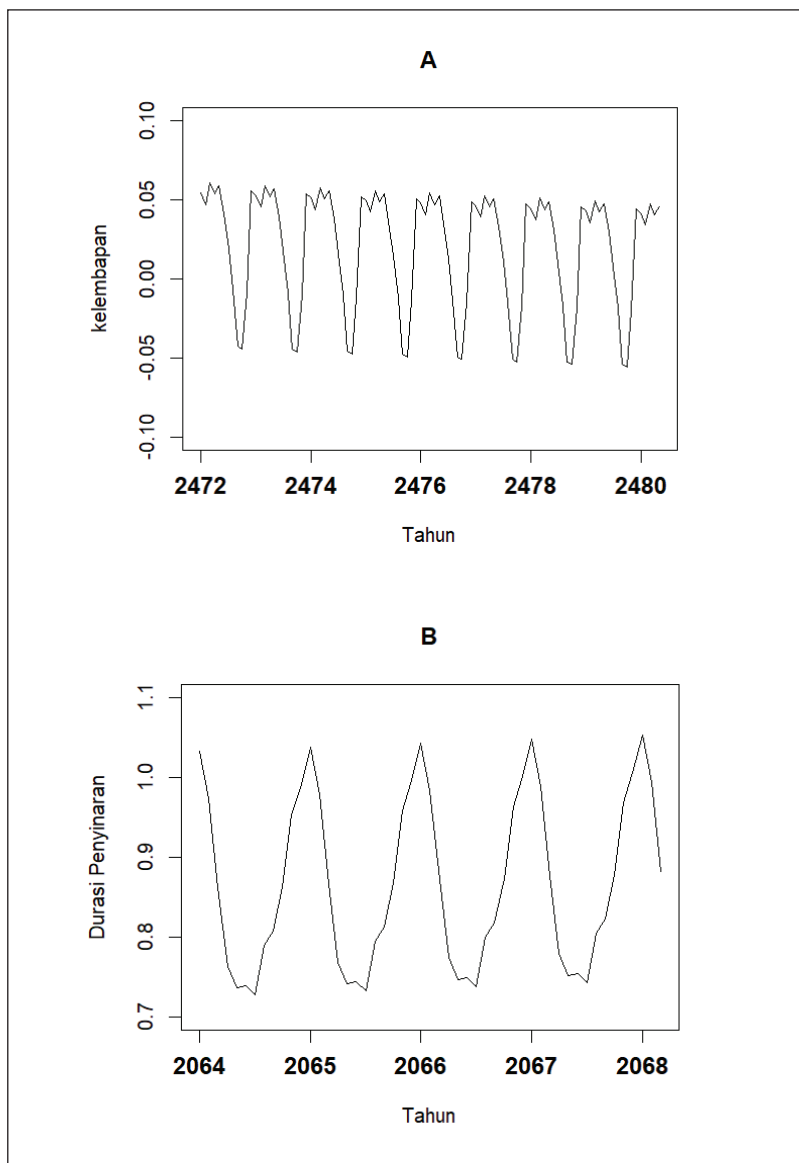
Prediksi iklim dan cuaca dilakukan dengan menggunakan teknik dinamis, atau menggunakan teknik statistik. Untuk prediksi iklim dan cuaca berbasis statistika, sering kali bersifat *autoregressive*, seperti *autoregressive integrated moving average* (McClymont et al., 2023), regresi linear (Sridhara et al., 2023), dan lainnya. Sebagai sebuah proses statistik, prediksi melibatkan banyak bilangan. Konsep bilangan pada matematika merupakan konsep dasar yang penting dalam pemodelan dan prediksi iklim. Dalam hal ini, beberapa parameter cuaca dan iklim memiliki karakteristik sendiri yang tidak sama dengan parameter lainnya. Mengabaikan karakteristik yang melekat pada masing-masing parameter sering kali berakibat pada hasil prediksi yang tidak logis. Sebagai perbandingan, parameter suhu sebuah benda, dapat berada di bawah 0°C , seperti es, atau di atas 400°C , seperti suhu lava yang mana hal tersebut sangat tergantung pada sifat benda. Inti matahari dapat mencapai suhu mencapai 15 juta derajat celsius.

Berbeda dengan konsep suhu, nilai parameter kelembapan relatif hanya berkisar antara 0%–100%. Ini terkait dengan pengertiannya sendiri sebagai sebuah hasil perbandingan antara jumlah uap air di dalam udara (kelembapan mutlak) dengan jumlah uap air maksimum yang dapat dikandung dalam udara tersebut. Demikian juga dengan nilai parameter durasi penyinaran. Karena keterbatasan rentang nilai tersebut, perlu kehati-hatian dalam proses pemodelan dan prediksinya. Dalam praktiknya, terkadang proses pemodelan dan prediksi iklim membutuhkan metode transformasi data untuk membantu menjaga nilai parameter sesuai dengan kaidahnya. Namun,

sesuai dengan sifat sistem iklim yang kompleks, penggunaan metode tersebut, tidak berarti memberi implikasi positif pada peningkatan akurasi dan resolusi hasil model atau prediksi. Dalam bab ini, penulis akan menunjukkan kompleksitas sistem prediksi di mana penggunaan metode transformasi hanya dapat memberi kontribusi untuk menjaga konsistensi nilai hasil prediksi sesuai dengan kaidah dan kriteria parameter iklim.

Dalam demonstrasi transformasi data ini, penulis menggunakan kasus kelembapan dan durasi penyinaran di Kota Bandung. Penyajian data dan analisis prediksinya ditunjukkan pada bagian permodelan (Subbab D) dengan menggunakan data dari tahun 2000 sampai 2022 untuk dapat menghasilkan model prediksi untuk beberapa tahun ke depan. Prediksi dilakukan menggunakan *time series linear model* (TSLM) untuk mencakup kondisi linieritas antarperiode dan fungsi *autoregressive* (Paolella, 2018). Prediksi dilakukan sampai tahun 2480 untuk kelembapan dan durasi penyinaran sampai tahun 2068. Dasar pertimbangan dalam penentuan periode prediksi tersebut adalah pola kelembapan dan durasi penyinaran di Kota Bandung yang cenderung stagnan sehingga nilai yang tidak konsisten atau di luar kaidah, baru terlihat pada periode tersebut. Terdapat kemungkinan bahwa daerah lain yang memiliki tren lebih tajam maka ketidakkonsistenan mungkin akan terlihat dalam periode yang lebih pendek. Bab ini hanya menyajikan contoh kasus di mana muncul nilai-nilai hasil prediksi yang di luar kaidah dan bagaimana mengatasinya.

Pada Gambar 8.1 A, prediksi kelembapan menjadi di luar kaidah dengan memiliki nilai kurang dari 0. Prediksi durasi penyinaran juga akan memiliki nilai yang di luar kaidah dengan memiliki nilai yang melebihi 100% di Gambar 8.1 B. Meskipun prediksi yang dilakukan cukup jauh karena cuaca Kota Bandung tidak memiliki variansi ekstrem dan tren yang cenderung stagnan, prediksi hasil perhitungan menunjukkan adanya nilai kelembapan udara dan durasi penyinaran matahari di luar kaidah.



Sumber: Januardi (2024)

Gambar 8.1 Kesalahan Estimasi Prediksi pada Kelembapan (A) dan Durasi Penyinaran (B)

Bab ini dibagi menjadi beberapa subbab. Subbab B mengenalkan ke pembaca beberapa variabel yang ditemui pada prakiraan cuaca. Subbab C memberikan cara dari proses transformasi data, prediksinya, sampai ke uji akurasi. Dalam Subbab D disampaikan studi kasus transformasi data kelembapan dan durasi penyinaran Kota Bandung yang didasarkan pada data dari Badan Pusat Statistik (2022a, 2022b). Penjelasan lebih rinci terkait ketidakcocokan tanpa transformasi *log-odds* dijelaskan di Subbab E. Subbab F memberikan implikasi secara teoretis dan manajerial dalam penggunaan transformasi data untuk prediksi cuaca. Terakhir, Subbab G merangkum kajian analisis yang dilakukan.

B. Variabel-Variabel Kasus pada Prakiraan Cuaca

Kelembapan merupakan konsentrasi air yang teruapkan pada udara sekitar. Air yang teruapkan dalam kondisi gas tidak dapat dilihat oleh kasat mata. Tingginya kelembapan akan mengindikasikan potensi terjadinya presipitasi yang secara visual dapat terdeteksi dari adanya awan dan kabut (Korotcenkov, 2020). Estimasi kelembapan dapat diketahui melalui kelembapan absolut (*absolute humidity*) dan kelembapan relatif (*relative humidity*). Kelembapan absolut adalah jumlah massa atau volume dari air yang teruapkan per massa atau volume udara kering (gram per kilogram). Kelembapan absolut memiliki tipe data yang terdistribusi normal, sedangkan kelembapan relatif menunjukkan perbandingan jumlah uap air di udara aktual dengan jumlah uap air maksimum yang bisa dikandung oleh udara pada suhu yang sama. Kelembapan relatif ini merupakan salah satu indikator penting untuk memprediksikan cuaca. Cuaca akan tergolong kering apabila kelembapan relatifnya cenderung rendah dan cuaca tergolong basah bila kelembapan relatifnya tinggi. Persamaan untuk mengetahui nilai kelembapan relatif (φ) menurut Rumble (2017) adalah sebagai berikut.

$$\varphi = \frac{p}{p_s} \times 100\% \quad (8.1)$$

Kelembapan relatif (φ) adalah rasio dari tekanan parsial dari uap air di udara (P) dengan tekanan uap jenuh pada udara pada suhu yang tertentu (p_s). Nilai P mengilustrasikan kondisi nyata (*real*) jumlah air yang teruapkan dan nilai p_s adalah potensi penguapan air secara keseluruhan yang tergantung pada kondisi suhunya. Kondisi lingkungan dengan kelembapan relatif yang rendah merupakan implikasi dari rendahnya tekanan parsial P . Selain itu, suhu yang ada akan menentukan jumlah uap air maksimal yang dapat ditampung di udara. Makin tinggi suhu, volume udara meningkat, yang berarti uap air yang dapat ditampung menjadi lebih banyak pula, sedangkan daerah yang memiliki kelembapan relatif tinggi mengindikasikan tingginya tekanan parsial, yang pada titik kritis tertentu mendorong terjadinya proses termodinamika dengan terjadinya perubahan fasa dengan asumsi tidak ada perubahan suhu. Naiknya tekanan parsial uap air sampai melewati titik jenuh berpotensi mengalami kondensasi, uap air menjadi droplet yang membentuk kabut atau awan. Dengan adanya proses penggabungan droplet menjadi tetes air yang lebih besar dan berkecepatan jatuh yang lebih tinggi maka tercipta presipitasi.

Nilai kelembapan relatif menggunakan satuan persentase untuk menggambarkan rasio antara tekanan parsial (P) dengan tekanan uap jenuh air di lingkungan (p_s). Saat bernilai kelembapan relatif bernilai 100%, tekanan parsial akan sama dengan tekanan uap jenuh. Pada titik ini, udara tidak dapat menampung air lagi sehingga tidak akan ada lagi air yang dapat teruapkan. Air pada lingkungan tersebut sudah mencapai titik embun (*dew point*) (Korotcenkov, 2018). Nilai dari kelembapan relatif dapat melebihi 100% yang akan mengakibatkan pembentukan formasi awan tertentu dan terjadinya hujan. Akan tetapi, nilai dari kelembapan relatif tidak mungkin bernilai di bawah 0% karena nilai 0% itu sendiri telah mengartikan tidak ada kesediaan air sama sekali di udara. Pada kenyataannya, tidak ada bagian di bumi ini yang memiliki nilai kelembapan relatif 0% karena 71% dari bumi adalah air.

Ketergantungan kelembapan relatif pada suhu menjadikan kelembapan relatif memiliki sifat yang tidak pasti (*uncertainty*). Meskipun

kandungan air dalam udara tetap, kelembapan relatif dapat berubah jika terjadi perubahan suhu. Pengolahan data kelembapan relatif dapat berbentuk *univariate* di mana pola pemunculan angkanya hanya tergantung dari pola deret waktu (*time series*) dan pola musiman (Nagaraj & Kumar, 2023).

Adapun durasi penyinaran adalah indikator dalam klimatologi untuk mengukur durasi penyinaran matahari dalam sebuah periode (satu hari, bulan, atau tahun) pada lokasi tertentu. Variabel ini juga dapat menjadi indikator keberadaan awan dari sebuah lokasi. Bentuk variabel ini merupakan rasio durasi penyinaran sesungguhnya terhadap maksimum durasi penyinaran di suatu daerah (Rahim et al., 2004). Pada aplikasinya, estimasi dari durasi penyinaran sangat bergantung dari sudut penyinaran dari matahari (*solar altitude*). Menurut Aydinli dan Krochmann (1987), estimasi durasi penyinaran ditunjukkan pada persamaan berikut.

$$\sigma_n = \left(1 - \frac{0.1}{\tan \gamma_s}\right) \times 100\% \quad \text{untuk } \gamma_s > 10^\circ \quad (8.2)$$

$$\sigma_n = 2.5 \tan \gamma_s \times 100\% \quad \text{untuk } \gamma_s \leq 10^\circ \quad (8.3)$$

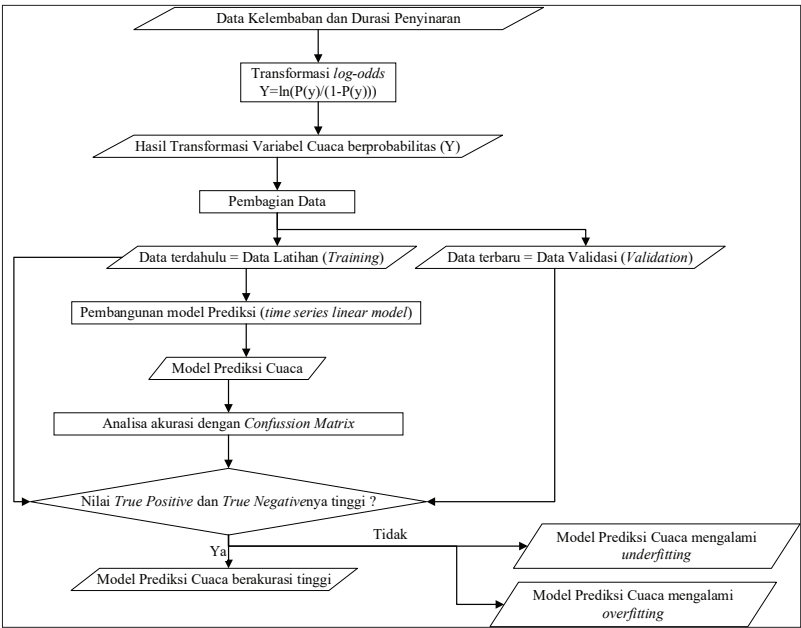
Nilai σ_n memiliki sifat tidak berdimensi (*dimensionless*), tetapi karena analisisnya menyangkut rasio sehingga dapat diklasifikasikan dengan persentase (Johansson, 2010). Lama penyinaran (σ_n) cukup bergantung pada sudut penyinaran dari matahari (γ_s). Jika matahari berada di tegak lurus pada suatu daerah ($\gamma_s \approx 90^\circ$) maka persentase penyinaran matahari akan mendekati 100%. Hal ini yang menyebabkan daerah-daerah ekuator di bumi mendapat penyinaran matahari paling tinggi. Akan tetapi, jika sudut penyinarannya matahari sejajar ($\gamma_s \approx 0^\circ$) maka persentase penyinaran matahari akan mendekati 0%. Hal ini terjadi untuk daerah di kutub selatan. Rendahnya penyinaran matahari menyebabkan suhu menjadi rendah dan dapat mendekati titik beku.

Lama penyinaran matahari sama seperti kelembapan memiliki sifat yang tidak pasti (*uncertain*) dan dapat diprediksi dengan pendekatan univariat menggunakan analisis deret waktu (Santos et al.,

2021). Durasi penyinaran matahari dengan probabilitas suatu daerah akan disinari oleh matahari secara terus menerus ($P(y) \approx 1$) atau memiliki probabilitas tidak disinari oleh matahari secara berkelanjutan ($1 - P(y) \approx 1$) (Maafi & Adane, 1998). Analisis statistika yang dapat dilakukan untuk variabel ini perlu dilakukan transformasi.

C. Proses Pengolahan dan Pengujian Data

Pada bagian ini, proses pengolahan data dari transformasi data, pengembangan model prediksi sampai pengujian data dilakukan. Beberapa proses akan dipertegas sebagai pembeda untuk pengolahan tanpa transformasi dan dengan transformasi, meski keduanya menggunakan model prediksi yang sama, yang dalam contoh ini menggunakan *time series linear model* (TSLM). Alur dari pengolahan data ditunjukkan dengan diagram alir pada Gambar 8.2.



Gambar 8.2 Diagram Alir dari Penngolahan Data Cuaca yang Disertai Transformasi Data

1. Transformasi *Log-Odds*

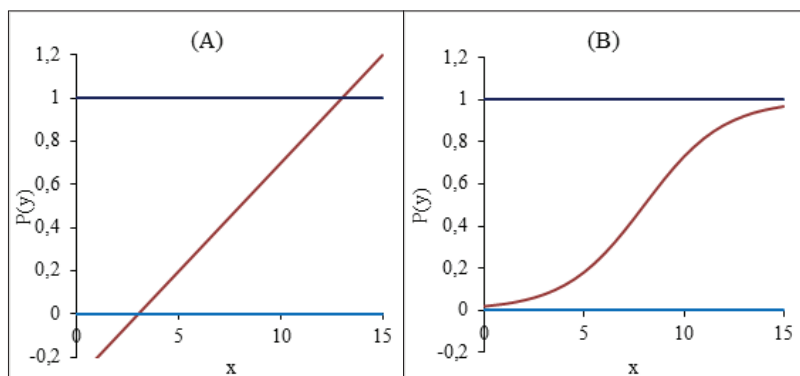
Teknik transformasi data yang digunakan untuk demonstrasi ini adalah fungsi *Log-odds*. Fungsi *log-odds* (*logit*) pada umumnya digunakan oleh pengolah data dan statistikawan untuk mengolah data yang memiliki batasan dan tidak memiliki distribusi normal. Apabila terdapat sebuah data yang memiliki dua jenis kelas (batas atas dan batas bawah), data dapat diklasifikasikan pada suatu kelas dengan probabilitas $P(y)$ atau terklasifikasi pada kelas lain dengan probabilitas $1 - P(y)$ tanpa distribusi normal. Data yang tidak terdistribusi normal, cenderung sulit untuk dianalisis atau diolah. Jika suatu data memiliki rentang nilai 0 sampai 1 [0,1], perlu dilakukan transformasi, salah satunya dengan menggunakan fungsi *log-odds* sehingga memiliki rentang nilai dari minus tak hingga sampai positif tak hingga $[-\infty, \infty]$ dan terdistribusi normal. Transformasi *log-odds* dilakukan dengan fungsi berikut ini menurut Hosmer et al. (2013).

$$Y = \ln \left(\frac{P(y)}{1 - P(y)} \right) \quad (8.4)$$

Probabilitas dari kejadian x ($P(x)$) dapat dijadikan sebuah variabel (Y) dengan logaritma natural. Syaratnya adalah nilai $P(y)$ harus memiliki rentang $0 \leq P(y) \leq 1$. Probabilitas, rasio *odds*, dan *log-odds* bukan merupakan hal yang sama. Ketiga hal ini menggambarkan proses transformasi dari variabel yang sudah ditransformasikan menjadi variabel yang berbeda. Probabilitas menggambarkan kemungkinan sebuah *odds* akan terjadi dalam unit persentase. Rasio *odds* adalah perbandingan antara dua kemungkinan kejadian (*odds*) jika memiliki antara sebuah *odds* terjadi ($P(y)$) dan tidak terjadi ($1 - P(y)$). Kemudian, dengan fungsi *log-odds* kedua *odds* memiliki bias (variansi) dengan rentang $[-\infty, \infty]$ sehingga dapat diaplikasikan metode statistika dasar. Setelah diperoleh nilai yang sudah ditransformasikan (Y), variabel dapat dilakukan transformasi balik (*detransformation*) menggunakan fungsi ekponensial sebagai berikut.

$$P(y) = \frac{e^Y}{1 + e^Y} = \frac{1}{1 + e^{-Y}} \quad (8.5)$$

Esensi dari fungsi *log-odds* adalah rasio antara probabilitas sebuah kejadian x ($P(y)$) terhadap probabilitas selain kejadian x ($1 - P(y)$). Pemakaian fungsi ini membentuk fungsi *sigmoid* yang berguna untuk mencegah nilai probabilitas yang dihasilkan keluar dari rentang nilai 0 sampai 1 $[0,1]$. Secara grafik model, Gambar 8.2 menunjukkan perbedaan tanpa dan menggunakan transformasi *log-odds*. Berdasarkan hasil fungsi prediksi, bertambah atau berkurangnya variabel bebasnya akan membuat prediksi probabilitas bernilai di luar rentang (Gambar 8.3 A) dan membuat data model yang dihasilkan tidak mengikuti kaidah. Alhasil pengambilan keputusan mengikuti model tanpa transformasi akan sulit. Berbeda halnya dengan proses transformasi *log-odds* (Gambar 8.3 B) yang dapat mencegah sebuah prediksi di luar rentang $[0,1]$ karena penggunaan dari fungsi *sigmoid*.



Sumber: Januardi (2024)

Gambar 8.3 Perbedaan Penggunaan Tanpa (A) dan dengan (B) Transformasi *Log-odds*

Dengan fungsi transformasi *log-odds*, sebuah kemungkinan (*odds*) dapat dimodelkan mengikuti variabel tertentu. Jika dalam sebuah probabilitas kejadian pasti terjadi, $P(y)=1$ maka kejadian lain tidak akan terjadi $1 - P(y) = 0$. Begitu pula untuk hal sebaliknya. Dengan fungsi logaritmik dari *log-odds* (persamaan 8.4), variabel yang ditransformasikan akan memiliki nilai dengan rentang tak hingga

$[-\infty, \infty]$ sehingga dapat diaplikasi pada statistika dasar. Pembuktiannya dapat dilihat pada persamaan berikut.

$$\ln\left(\frac{P(y) = 1}{1 - P(y) = 0}\right) = \ln\left(\frac{1}{0}\right) = \infty \quad (8.6)$$

$$\ln\left(\frac{P(y) = 0}{1 - P(y) = 1}\right) = \ln\left(\frac{0}{1}\right) = -\infty \quad (8.7)$$

Sebuah kemungkinan (*odds*) memiliki probabilitas dengan rentang 0–1. Nilai 1 menggambarkan sebuah *odds* pasti terjadi dan 0 menggambarkan sebuah *odds* pasti tidak terjadi. Logaritma *log-odds* adalah fungsi transformasi dari variabel dengan batasan menjadi variabel kontinu yang dapat mencapai tak terhingga $[-\infty, \infty]$. Data-data kemungkinan ini diperoleh secara statistik dari data historis. Sebagai contoh, probabilitas sebuah daerah akan hujan bernilai 80% maka probabilitas tidak terjadi hujannya adalah 20%. Jika menggunakan *log-odds* maka nilai $\ln(80\%/20\%)$ akan menjadi 1.38. Nilai *log-odds* ini yang dapat diaplikasikan pada model statistika dasar.

Penggunaan transformasi *log-odds* adalah cara mendasar untuk mengaplikasikan variabel yang memiliki batasan (tidak tak terhingga) pada model statistika dengan proses yang didapat diikuti dalam setiap tahapannya. Selain transformasi *log-odds*, prediksi variabel dapat menggunakan metode *machine learning*, seperti *support vector machine*, pohon klasifikasi (*classification tree*), atau jaringan saraf tiruan (*neural network*). Namun, menurut Rudin (2019), prediksi variabel secara langsung menggunakan *machine learning* akan mengindikasikan kejadian *black box*, di mana hasil prediksi yang dibentuk tidak dapat dimengerti proses pembentukannya (masih bersifat abstrak). Keabstrakan dari sebuah model akan mengakibatkan sulitnya implementasi dari model prediksi untuk proses pengambilan keputusan selanjutnya.

2. Time Series Linear Model

Metode *time series linear model* (TSLM) merupakan pengembangan dari model regresi linear berganda (*multiple linear regression*). Metode TSLM ini memiliki parameter khusus pada variabel waktu ke- t dan variabel pada waktu sebelumnya. Penggunaan variabel pada waktu sebelumnya adalah fungsi autoregresif (*autoregressive*) untuk melihat nilai musimannya (*seasonality*). Secara permodelan matematika menurut Paolella (2018), model TSLM ditunjukkan sebagai berikut.

$$f(x) = \beta_0 + \beta_1 t + \beta_2 x_{t-1} + \beta_2 x_{t-2} + \dots \quad (8.8)$$

Di mana fungsi dari model TSLM memiliki parameter sebagai *slope* untuk variabel waktu (t) dan variabel univariat terdahulu (x_{t-i}). Penggunaan variabel terdahulu digunakan untuk menjelaskan kemungkinan adanya serial korelasi (*autocorrelation*) pada kestokastikan data deret waktu. Pembentukan model masih menggunakan metode estimasi galat kuadrat terkecil (*least square estimator*). Pada artikel ini, pemakaian metode ini menggunakan fungsi *tslm* pada pemrograman R (Hyndman & Athanasopoulos, 2018).

3. Akurasi Model dengan Confusion Matrix

Prediksi deret waktu memiliki penilaian kinerja model menggunakan analisis galat, seperti *mean square error* (MSE), *mean absolute error* (MAE), *mean absolute percentage error* (MAPE), *root mean square error* (RMSE), dan lainnya. Namun, untuk mengevaluasi hasil kinerja transformasi *log-odds* digunakan metode *confusion matrix* (Hastie et al., 2001). Matriks ini menggunakan analisis akurasi apakah probabilitas yang dihasilkan memenuhi frekuensi tertentu. Persamaan untuk menilai akurasi dari model menurut Sammut dan Webb (2010) adalah sebagai berikut.

$$\text{Akurasi}(\%) = \frac{TP_n + TN_n}{n} = \frac{TP_n + TN_n}{TP_n + TN_n + FP_n + FN_n} \quad (8.9)$$

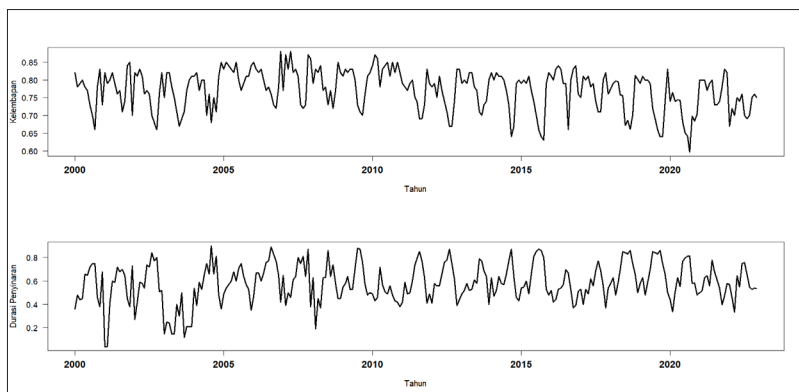
Persamaan akurasi hanya dapat digunakan untuk data yang memiliki batas atas dan batas bawah. Penetapan *threshold* diperlukan untuk mengklasifikasikan data kepada nilai positif (*positive*) atau negatif (*negative*). Keterangan TP_n (*true positive*) adalah kumpulan hasil prediksi model di mana data yang keluar sesuai masuk ke dalam suatu kelas, sedangkan keterangan TN_n (*True Negative*) adalah kumpulan prediksi di mana data yang keluar sesuai masuk ke dalam kelas lain. Nilai merujuk pada jumlah hasil prediksi yang keluar dari model. Tingginya frekuensi TP_n dan TN_n menunjukkan kalau model berhasil untuk memprediksikan sebuah kejadian. Keterangan FP_n (*False Positive*) dan FN_n (*False Negative*) menunjukkan kalau prediksi merujuk ke suatu kelas, padahal keadaan yang sesuai adalah pada kelas lain. Tingginya frekuensi nilai FP_n dan FN_n menunjukkan rendahnya kemampuan model prediksi untuk memprediksikan data.

D. Studi Kasus Kota Bandung

Untuk memudahkan aplikasi dan analisis dari model prediktif variabel pada cuaca, subbab ini menggunakan data cuaca dari Kota Bandung pada tahun 2000 sampai tahun 2022 menurut data dari Badan Pusat Statistik (2022a, 2022b). Kemudian data dipisah menjadi data latihan (*training data*) dan data validasi (*validation data*) menggunakan prinsip *pareto* (Panesar, 2021). Data dari tahun 2000–2021 digunakan untuk membangun model prediksi (data latihan) dan data tahun 2022 dipakai untuk validasi hasil prediksi. Pembuatan model akan dibagi menjadi dua, yaitu model tanpa dan dengan proses transformasi *log-odds*.

1. Data Cuaca Kota Bandung

Data yang dipakai untuk prediksi cuaca adalah data kelembapan dan durasi penyinaran dari Kota Bandung di tahun 2000–2022 (Badan Pusat Statistik, 2022a, 2022b). Penampakan dan statistika deskriptif dari data cuaca Kota Bandung ditunjukkan pada Gambar 8.3.



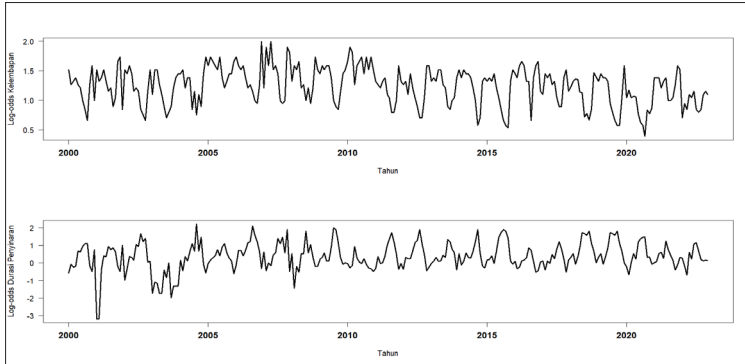
Sumber: BPS (2022a, 2022b)

Gambar 8.4 Data Kelembapan dan Durasi Penyinaran Kota Bandung dari Tahun 2000–2022

Tabel 8.1 Statistika Deskriptif Data Kelembapan dan Durasi Penyinaran Kota Bandung dari Tahun 2000–2022

Parameter	Kelembapan	Durasi Penyinaran
Minimum	0,5969	0,04
Median	0,79	0,58
Rata-rata	0,772	0,5825
Maksimum	0,88	0,9

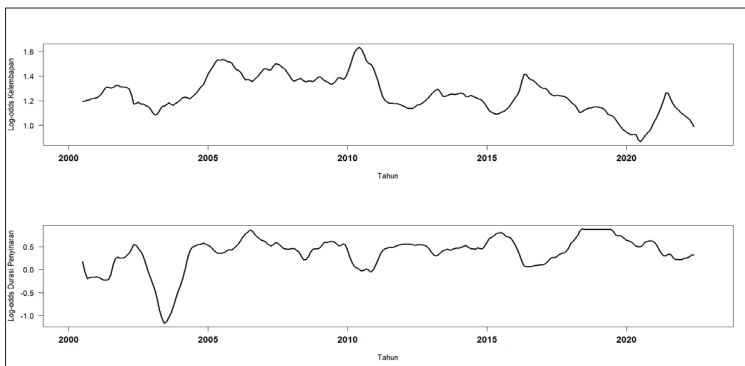
Apabila dilakukan analisis statistika deskriptif, pola data durasi penyinaran memiliki rentang data yang lebih tinggi dibandingkan data kelembapan. Hal ini dapat dilihat dari perbedaan nilai maksimal dan minimalnya pada Tabel 8.1. Data yang ada dilanjutkan untuk transformasikan menggunakan fungsi *log-odds* yang mengikuti persamaan 8.4. Gambar 8.4 menunjukkan hasil transformasi *log-odds*nya.



Sumber: BPS (2022a, 2022b)

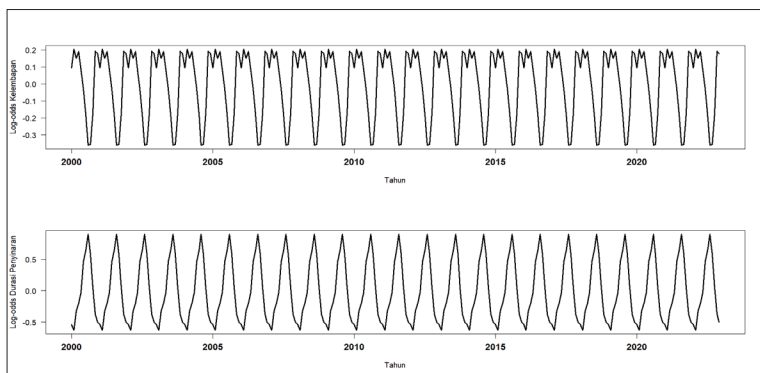
Gambar 8.5 Transformasi *Log-odds* dari Data Kelembapan dan Durasi Penyiaran Kota Bandung dari Tahun 2000–2022

Berdasarkan perbedaan data yang belum ditransformasi (Gambar 8.3) dan yang sudah ditransformasi (Gambar 8.4), tidak menunjukkan pola perbedaan, hanya menunjukkan perbedaan pada nilai rentang dari data. Kemudian, data diurai menggunakan fungsi (*decompose*) pada pemrograman R. Penguraian data deret waktu (*decomposing*) dilakukan untuk menganalisis tren (kecenderungan naik dan turun) dan musiman (pola spesifik setiap musim) (Chatfield & Xing, 2019). Hasilnya adalah sebagai berikut.



Sumber: Januardi (2024)

Gambar 8.6 Analisis Tren dari *Log-odds* Kelembapan dan Durasi Penyiaran Kota Bandung dari Tahun 2000–2022



Sumber: Januari (2024)

Gambar 8.7 Analisis Musiman (*Seasonality*) dari *Log-odds* Kelembapan dan Durasi Penyinaran Kota Bandung dari Tahun 2000–2022

Hasil dari Gambar 8.5 mengindikasikan kondisi stokastik dari *log-odds* kelembapan dan durasi penyinaran. Artinya kedua data tidak menunjukkan kecenderungan naik atau turun. Namun hasil analisis pada Gambar 8.6 menunjukkan bahwa *log-odds* dari kelembapan dan durasi penyinaran memiliki sebuah pola musiman. Setiap lima tahun sekali, Kota Bandung akan mengalami enam kali nilai kelembapan yang tinggi dan lima kali nilai durasi penyinaran yang tinggi. Hal ini mendukung bahwa kedua jenis data merupakan deret waktu dengan pola musiman.

2. Permodelan Prediksi

Setelah data kelembapan dan durasi penyinaran dari Kota Bandung dianalisis, selanjutnya data digunakan untuk membangun model prediksi. Model prediksi yang digunakan adalah *time series linear model* (TSLM) mengikuti persamaan 8.8. Model prediksi kelembapan dan durasi penyinaran dibentuk dengan data yang tanpa transformasi dan dengan transformasi *log-odds*. Model prediksi yang dibangun menggunakan data dari tahun 2000 sampai 2021. Hasil prediksinya akan dilakukan uji akurasi dengan data dari tahun 2022 mengikuti prinsip *pareto*. Hasil estimasi statistik dan hasil pengujian untuk data kelembapan adalah sebagai berikut.

Tabel 8.2 Estimasi Statistik dari Model TSLM Kelembapan

	Tanpa Transformasi ($P(y)$)		Dengan Transformasi ($\ln(\frac{P(y)}{1-P(y)})$)	
	Estimasi	P-value	Estimasi	P-value
Model		<0,0001		<0,0001
Intercept	0,8146*	<0,0001	1,4766*	<0,0001
Tren (t)	-0,0001*	<0,0001	-0,0008*	<0,0001
Musiman		<0,0001		<0,0001
Bulan 2	0,0138	0,2634	0,0889	0,2079
Bulan 3	0,0068	0,5833	0,0428	0,5442
Bulan 4	0,0120	0,3296	0,0771	0,2748
Bulan 5	-0,0055	0,6578	-0,0355	0,6149
Bulan 6	-0,0272*	0,03	-0,1540*	0,0297
Bulan 7	-0,0533*	<0,0001	-0,2952*	<0,0001
Bulan 8	-0,0884*	<0,0001	-0,4791*	<0,0001
Bulan 9	-0,0901*	<0,0001	-0,4772*	<0,0001
Bulan 10	-0,0544*	<0,0001	-0,2937*	<0,0001
Bulan 11	0,0097	0,4328	0,0717	0,3101
Bulan 12	0,0071	0,5658	0,0576	0,4144

Keterangan: *Signifikan pada 95%

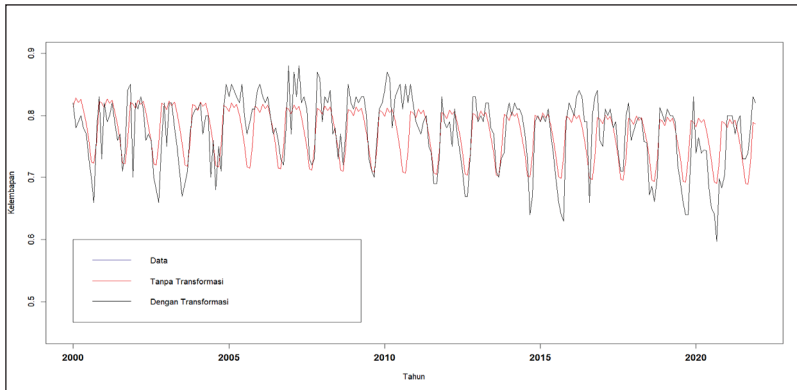
Tabel 8.2 menunjukkan estimasi model statistika prediktif dari kelembapan tanpa dan dengan transformasi *log-odds*. Kedua model menunjukkan hasil signifikansi yang sama. Kedua model memiliki signifikansi musiman pada bulan Juni sampai Oktober untuk prediksi deret waktu kelembapan. Variabel bulan Januari termasuk ke dalam nilai dasar atau *intercept*. Berdasar dari penampakan Tabel 8.2, baik dengan transformasi maupun tanpa transformasi tidak menunjukkan perbedaan pada nilai signifikansi. Tabel selanjutnya menunjukkan hasil estimasi statistik dari data durasi penyinaran.

Tabel 8.3 Estimasi Statistik dari Model TSLM Durasi Penyinaran

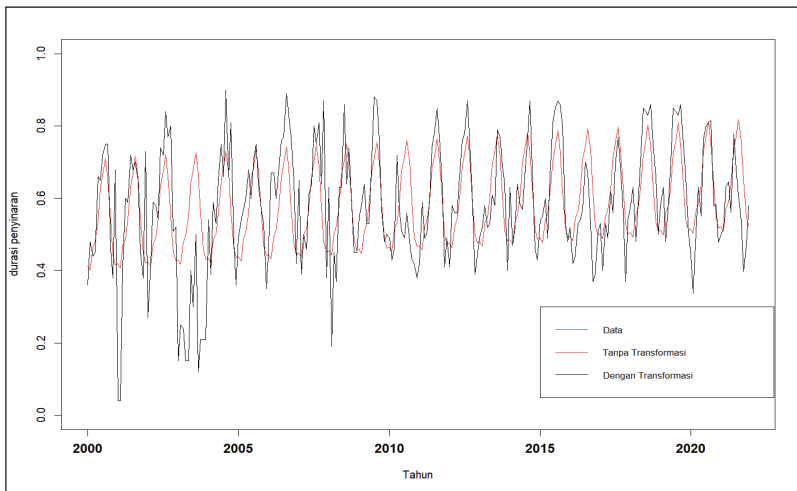
	Tanpa Transformasi ($P(y)$)		Dengan Transformasi ($\ln(\frac{P(y)}{1-P(y)})$)	
	Estimasi	P-value	Estimasi	P-value
Model		<0,0001		<0,0001
Intercept	0,4135*	<0,0001	-0,4818*	0,0015
Tren	0,0004*	<0,0001	0,0022*	<0,0001
Musiman		<0,0001		<0,0001
Bulan 2	-0,0123	0,7527	-0,0457	0,8110
Bulan 3	0,0488	0,2106	0,2704	0,1578
Bulan 4	0,0666	0,0880	0,3335	0,0817
Bulan 5	0,1210*	0,0021	0,5583*	0,0038
Bulan 6	0,2110*	<0,0001	1,0027*	<0,0001
Bulan 7	0,2480*	<0,0001	1,2005*	<0,0001
Bulan 8	0,2942*	<0,0001	1,4600*	<0,0001
Bulan 9	0,2347*	<0,0001	1,1340*	<0,0001
Bulan 10	0,1231*	0,0017	0,6041*	0,0017
Bulan 11	0,0249	0,5221	0,1864	0,3299
Bulan 12	-0,0026	0,9462	0,0573	0,7644

Keterangan: *Signifikan pada 95%

Tabel 8.3 adalah model statistika prediktif dari durasi penyinaran yang tanpa dan dengan transformasi *log-odds*. Sama seperti Tabel 8.2, signifikansi kedua model memiliki hasil yang sama. Berbeda dari kelembapan, model prediksi durasi penyinaran hanya signifikan pada bulan Mei sampai Oktober. Setelah model prediksi dibangun untuk variabel kelembapan dan durasi penyinaran, model diteruskan untuk diuji akurasiya menggunakan persamaan 8.9. Akurasi model dibedakan untuk data latihan (*data training*) dan validasi (*data validation*).



Gambar 8.8 Analisis Data Latihan Kelembapan dari Kota Bandung Tahun 2000–2021

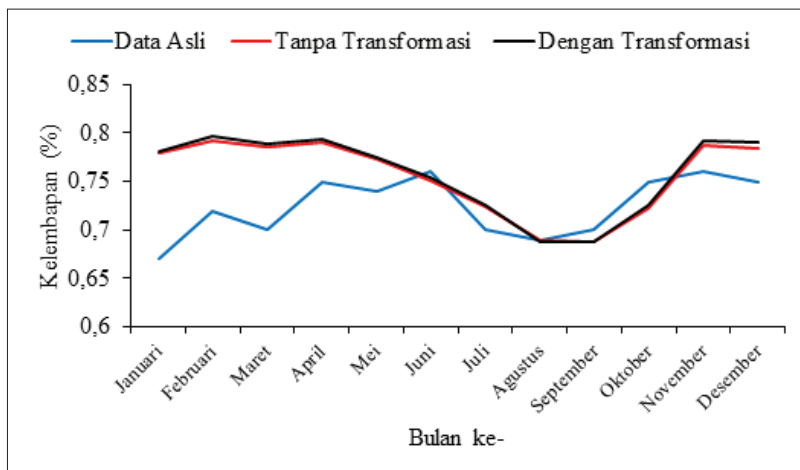


Sumber: Januardi (2024)

Gambar 8.9 Analisis Data Latihan Durasi Penyiaran dari Kota Bandung Tahun 2000–2021

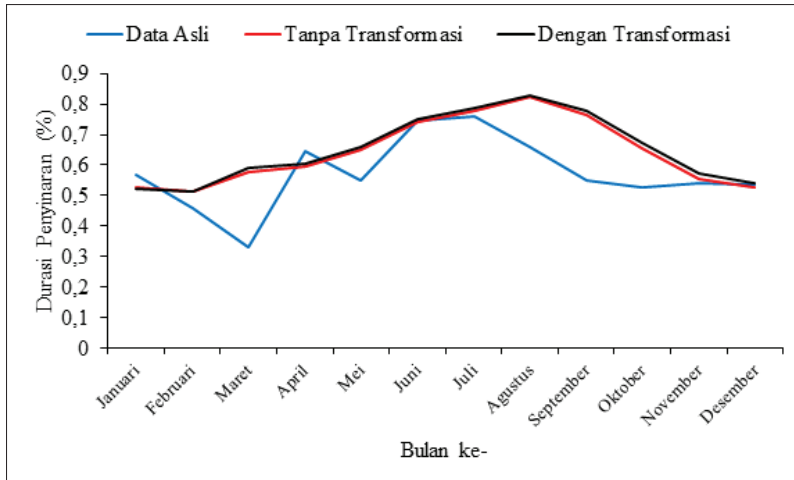
Apabila hasil prediksi kelembapan dan durasi penyiaran dari model tanpa transformasi dan dengan transformasi disejajarkan dengan data latihan (*training*) maka akan dihasilkan grafik, seperti yang tampil di Gambar 8.7 dan Gambar 8.8. Data asli dari tahun 2000–2021

ditunjukkan pada garis biru, model prediksi tanpa transformasi dengan garis merah, dan model prediksi dengan transformasi dengan garis berwarna hitam. Secara penampakan, model tanpa transformasi memiliki model yang kurang cocok terhadap data kelembapan, sedangkan model yang sudah ditransformasi *log-odds* memiliki kecocokan data yang sangat tinggi. Hal ini dibuktikan dengan garis biru (data awal) yang tidak terlihat karena tertutupi oleh garis hitam dari model prediksi yang tertransformasi. Indikasi model yang sangat cocok tentunya juga tidak terlalu bagus pada permodelan *machine learning*. Hasil model dari data latihan yang terlalu cocok terhadap datanya, akan mengindikasikan kejadian *overfitting*. Kejadian *overfitting* merupakan tingginya kecocokan model prediksi hasil data latihan yang mengakibatkan model tidak dapat memprediksikan data validasi sebagai data di luar latihan (Dobin, 2010). Pembandingan data validasi dengan hasil prediksi ditunjukkan pada Gambar 8.10.



Sumber: Januardi (2024)

Gambar 8.10 Analisis Data Validasi Kelembapan dari Kota Bandung Tahun 2022



Sumber: Januari(2024)

Gambar 8.11 Analisis Data Validasi Durasi Penyinaran dari Kota Bandung Tahun 2022

Pernyataan *overfitting* dari hasil data latihan dapat dilihat pada hasil data validasi di Gambar 8.9 dan 8.10. Pada Gambar 8.9, hasil prediksi baik tanpa dan dengan transformasi menunjukkan hasil yang cukup menyimpang pada rentang waktu Januari–April 2022. Namun model dapat mengikuti pola data validasi di saat waktu Mei–Agustus 2022. Hasil prediksi untuk tahun 2022 pada data kelembapan dan durasi penyinaran memiliki data yang lebih halus. Artinya perubahan data setiap bulannya tidak berubah secara tajam. Meski demikian, secara kasat mata, pola grafik dari hasil prediksi tanpa ataupun dengan transformasi memiliki pola yang sama dan hampir mengikuti pola dari data di tahun 2022. Sehingga model prediksi yang dihasilkan dapat digunakan untuk memprediksikan variabel cuaca. Untuk hasil yang lebih akurat, estimasi akurasi data latihan dan validasinya dilakukan menggunakan persamaan 8.9 dan ditunjukkan pada Tabel 8.4.

Tabel 8.4 Akurasi Model Kelembapan

	Tanpa Transformasi ($P(y)$)		Dengan Transformasi ($\ln(\frac{P(y)}{1-P(y)})$)	
	Data Latihan (Data Training)	Data Validasi (Data Validation)	Data Latihan (Data Training)	Data Validasi (Data Validation)
True Positive (TP)	25,00%	75,00%	34,09%	75,00%
False Positive (FP)	9,47%	8,33%	0,00%	8,33%
False Negative (FN)	18,56%	8,33%	9,47%	8,33%
True Negative (TN)	46,97%	8,33%	56,44%	8,33%
Akurasi	71,97%	83,33%	90,53%	83,33%

Tabel 8.5 Akurasi Model Durasi Penyinaran

	Tanpa Transformasi ($P(y)$)		Dengan Transformasi ($\ln(\frac{P(y)}{1-P(y)})$)	
	Data Latihan (Data Training)	Data Validasi (Data Validation)	Data Latihan (Data Training)	Data Validasi (Data Validation)
True Positive (TP)	45,45%	66,67%	43,56%	75,00%
False Positive (FP)	8,71%	16,67%	0,00%	8,33%
False Negative (FN)	14,02%	0,00%	2,27%	8,33%
True Negative (TN)	31,82%	16,67%	54,17%	8,33%
Akurasi	77,27%	83,33%	97,73%	83,33%

Pada model tanpa transformasi untuk prediksi kelembapan dan durasi penyinaran di Tabel 8.4 dan 8.5, prediksi data latihan (*training*) memiliki nilai akurasi yang lebih rendah dibandingkan akurasi data validasi. Hasilnya model menunjukkan hasil yang cukup bagus di mana kemampuan model TSLM dapat memprediksikan data-data di luar analisis (*outlier*), sedangkan pada model dengan transformasi, hasil data latihan (*training*) memiliki nilai akurasi yang lebih tinggi dibandingkan data validasi. Indikasinya adalah transformasi *log-odds* memperkecil nilai variansi dari data. Alhasil, kemampuannya untuk memprediksikan data *outlier* tidak terlalu bagus. Hal ini mungkin

akan mengindikasikan kondisi *overfitting*. Akan tetapi, apabila variansi data tidak dipertimbangkan (pada model tanpa transformasi), indikasi kesalahan prediksi dapat terjadi untuk data yang memiliki variansi cukup tinggi. Hal ini dibuktikan pada Gambar 8.1.

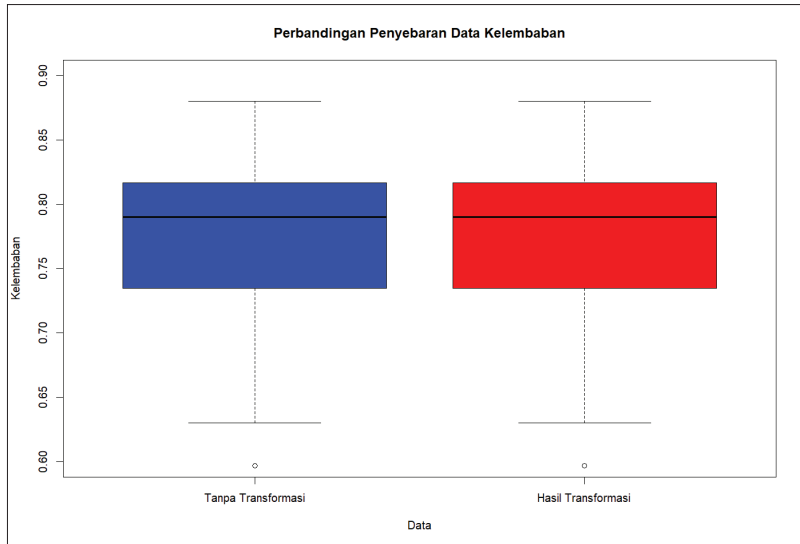
Dari hasil penilaian keakuratan data latihan dan validasi dari model yang tanpa dan dengan transformasi *log-odds*, beberapa hal menarik ditemukan pada data kelembapan dan durasi penyinaran. Pada data latihan, akurasi model dengan transformasi *log-odds* menunjukkan keakuratan yang cukup tinggi dibanding dengan model yang tanpa transformasi. Hal ini membuktikan bahwa proses transformasi akan mempertajam akurasi dari variabel untuk prediksi cuaca. Tabel 8.4 dan Tabel 8.5 juga menunjukkan bahwa model tanpa dan dengan transformasi menunjukkan keakuratan yang sama dari data validasi. Dengan demikian, apabila dipertimbangkan, dengan keakuratan data validasi yang sama dan memiliki keakuratan data latihan yang lebih tinggi, model dengan transformasi jauh lebih unggul dibanding dengan model tanpa proses transformasi *log-odds*.

E. Ketidakcocokan Model Tanpa Transformasi *Log-odds*

Setelah dilakukan aplikasi dan analisis prediksi dari variabel pada cuaca di subbab sebelumnya, subbab ini akan menjelaskan beberapa alasan perlunya proses transformasi *log-odds* untuk variabel tertentu di cuaca.

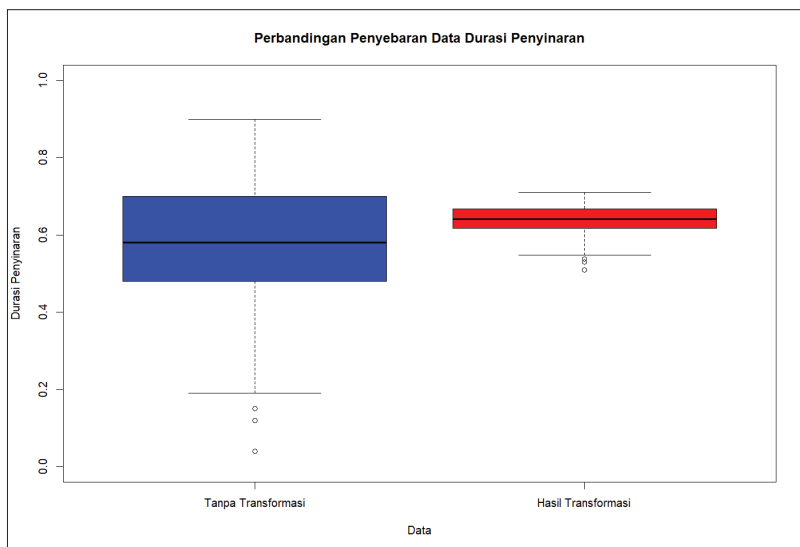
1. Penyebaran Data

Data yang langsung diaplikasikan model prediksi tanpa transformasi *log-odds* dapat memiliki pola penyebaran data yang cukup besar. Hal ini berkaitan dengan pola distribusi data yang cenderung tidak memiliki batas. Tidak seperti data berprobabilitas dengan rentang 0–1. Sebagai contoh, data dari kelembapan dan durasi penyinaran Kota Bandung dari tahun 2000–2022 dibentuk dalam *boxplot* berikut (Gambar 8.12 dan Gambar 8.13).



Sumber: Januardi (2024)

Gambar 8.12 *Boxplot* dari Kelembapan Kota Bandung Tahun 2000–2022



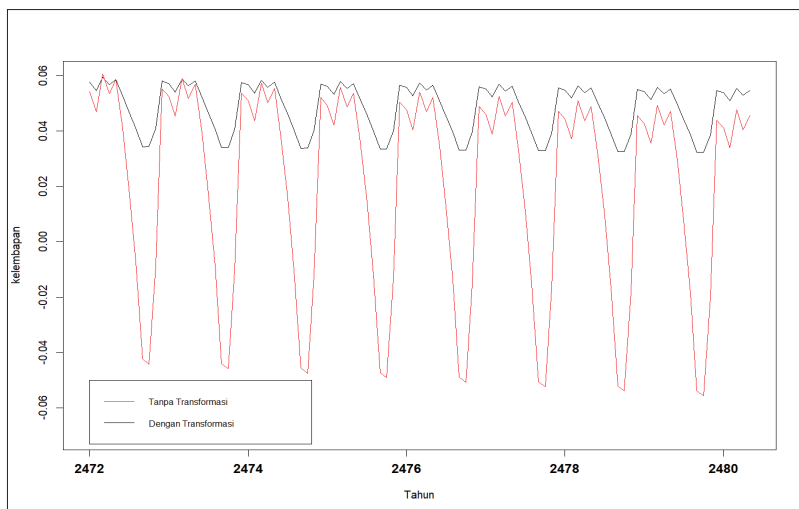
Sumber: Januardi (2024)

Gambar 8.13 *Boxplot* dari Durasi Penyinaran Kota Bandung Tahun 2000–2022

Grafik *boxplot* adalah alat statistika untuk menggambarkan penyebaran sebuah data dari nilai median, kuartil ke 2 dan 3, serta nilai minimum dan maksimalnya. Grafik ini cukup ampuh untuk melihat seberapa luas penyebaran data dari sampel (Black, 2023). Analisis *boxplot* yang dilakukan pada data kelembapan dan durasi penyinaran kota Bandung dari tahun 2000 sampai 2022 menghasilkan beberapa temuan. Data kelembapan baik yang tanpa dan dengan transformasi *log-odds* menunjukkan pola penyebaran data yang sama. Jarak data kelembapan memiliki rentang yang lebih pendek sehingga analisis hasilnya lebih mudah. Berbeda dari data kelembapan, data durasi penyinaran memiliki rentang data yang lebih besar. Alhasil perbedaan antara tanpa dan dengan transformasi *log-odds* cukup terlihat. Transformasi *log-odds* berguna untuk menjaga prediksi sebuah probabilitas untuk di rentang yang logis.

2. Prediksi yang Salah

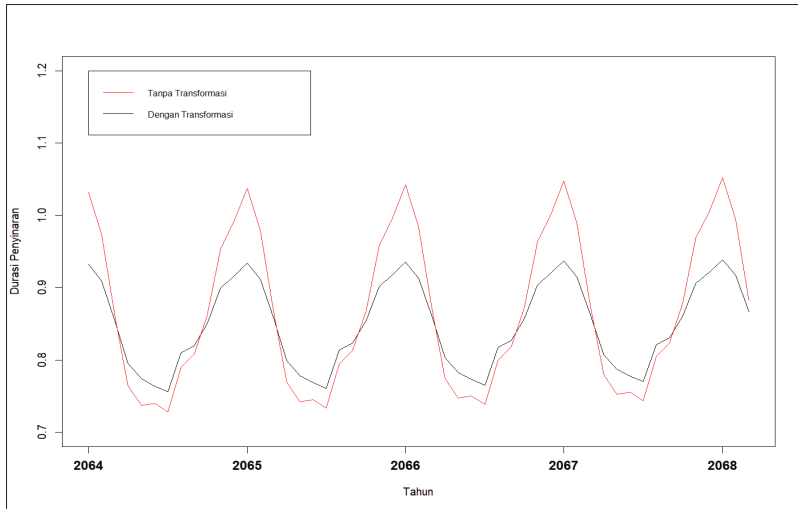
Prediksi deret waktu untuk jangka waktu yang lebih panjang (*longer sequence time series*) memiliki ketidakpastian (*uncertainty*) data yang cukup tinggi. Alhasil yang akan terjadi adalah nilai galat (*error*) yang dihasilkan juga akan lebih tinggi (Petropoulos et al., 2022). Kemungkinan linearitas data untuk waktu dalam jangka waktu ke depan dapat tidak terjadi. Hal ini mengindikasikan perlunya model *nonlinear* untuk aplikasi deret waktu. Namun, prediksi jangka waktu panjang tetap diperlukan untuk melihat pola data ke depannya. Karena kemungkinan hasil prediksi model dan data jangka waktu ke depan bernilai panjang memiliki nilai hampir 0, prediksi sebuah data deret waktu sangat dianjurkan untuk jangka waktu pendek. Setelah penjelasan untuk pola penyebaran suatu data dari model tanpa dan dengan transformasi *log-odds*, analisis dilanjutkan untuk prediksi cuaca. Kelembapan dan durasi penyinaran dilakukan prediksi untuk beberapa tahun ke depan di luar tahun 2022. Prediksi dilakukan untuk melihat kemungkinan sebuah data bernilai diluar rentang 0–1. Meskipun prediksi dilakukan untuk jangka waktu panjang, hasil prediksi hanya mengobservasi kemungkinan hasil prediksi di luar rentang variabel. Penampakan dari hasil prediksi kelembapan untuk tahun 2472 dilampirkan pada Gambar 8.14.



Sumber: Januardi (2024)

Gambar 8.14 Prediksi Kelembapan Kota Bandung Tahun 2472–2480

Prediksi kelembapan menemukan titik observasi yang menarik pada amatan di tahun 2472. Model tanpa transformasi akan memprediksikan kelembapan bernilai kurang dari 0. Nilai kelembapan relatif bernilai kurang dari 0 akan menyalahi kaidah matematika. Kelembapan yang bernilai kurang dari 0 menandakan sebuah daerah sangat kering (tanpa ada kadar air sama sekali) yang cukup mustahil terjadi di muka bumi. Hal ini menyebabkan tidak disarankan penggunaan variabel cuaca tanpa melakukan proses transformasi *log-odds*. Namun, keuntungan dari data kelembapan Kota Bandung adalah memiliki rentang nilai yang cukup rendah antara 60%–80% (menurut Tabel 8.1). Hal ini yang menyebabkan penyebaran data antara model tanpa dan dengan transformasi *log-odds* penyebarannya tidak berbeda jauh (Gambar 8.11). Apabila model kelembapan masih menggunakan model tanpa transformasi, lebih baik untuk aplikasi jangka waktu pendek.



Sumber: Januardi (2024)

Gambar 8.15 Prediksi Durasi Penyinaran Kota Bandung Tahun 2064–2068

Setelah variabel kelembapan, variabel cuaca yang dilakukan prediksi jangka waktu panjang adalah durasi penyinaran. Terlihat dari Gambar 8.15, jika prediksi durasi penyinaran diteruskan sampai tahun 2064, nilai durasi penyinaran akan melebihi 1. Hal ini juga menyebabkan data yang dihasilkan tidak memenuhi kaidah matematika (durasi penyinaran yang melebihi 100%). Mengikuti Persamaan 8.2 dan 8.3, beragamnya penetapan sudut penyinaran matahari juga membuat nilai durasi penyinaran tidak melebihi 100%. Penyinaran yang melebihi 100% menekankan bahwa sebuah daerah akan mengalami penyinaran secara terus menerus di atas cakupan penyinarannya. Hal ini cukup mustahil karena sebuah daerah akan mengalami siklus penyinaran yang tidak terus-menerus. Oleh karena itu, artikel ini menunjukkan bahwa perlunya proses transformasi *log-odds* terkhusus untuk variabel di cuaca. Hal ini bertujuan untuk mencegah hasil prediksi berada di luar rentang kaidah matematika yang sudah ditetapkan.

F. Implifikasi Teoretis dan Manajerial

Subbab ini akan menjelaskan berbagai implifikasi penggunaan transformasi *log-odds* untuk variabel pada cuaca secara teoretis dan manajerial. Secara teoretis, implifikasi dijelaskan menggunakan kaidah-kaidah matematika dan statistika, sedangkan secara manajerial, implifikasi dijelaskan kegunaannya untuk para praktisi, pemerhati, dan pemangku kebijakan ketahanan pangan.

1. Implifikasi Teoretis

Prediksi variabel cuaca, untuk variabel kelembapan dan durasi penyinaran, perlu memiliki batasan (batas atas batas bawah). Tanpa dilakukan pembatasan (*constraint effort*), hasil prediksi akan menghasilkan nilai yang di luar kaidah. Dalam proses pengambilan keputusan dan kebijakan, estimasi yang memenuhi kaidah matematika dan statistika sangat penting dan krusial. Bahasan ini perlu ditekankan untuk meyakinkan kepada pembaca agar melakukan proses transformasi *log-odds* terkhusus untuk variabel cuaca tertentu. Hal ini dibuktikan dari analisis pada Subbab E di mana hasil prediksi tanpa proses transformasi akan menghasilkan nilai yang di luar kaidah. Jika nilai yang dihasilkan di luar dari kaidah matematika dan statistika, proses selanjutannya akan sulit untuk dilakukan.

Sebuah penelitian dari Nagarajan dan Minu (2018) tentang penyemprotan tanaman pangan berbasis kelembapan menggunakan sensor dan prediksi kelembapan untuk penjadwalan penyemprotan saat kelembapan rendah. Apabila model tanpa transformasi diaplikasikan pada teknologi dari penelitian Nagarajan dan Minu (2018), kemungkinan sistem tidak berjalan akan tinggi karena dalam penetapan variabel kelembapan, kelembapan akan diatur dalam rentang tertentu. Jika prediksi kelembapan yang di luar kaidah ditetapkan pada sistem maka penyemprotan otomatis tidak terjadi atau terjadi namun dengan jumlah yang banyak (karena sistem membaca kelembapan yang sangat rendah di bawah 0). Ketidaksesuaian penyemprotan akan menurunkan kinerja dari pertumbuhan tanaman pangan.

Penelitian lain yang dilakukan oleh Saadon et al. (2021) mengembangkan pembukaan atap sebuah rumah kaca (*greenhouse*) untuk tanaman pangan yang cukup sensitif terhadap perubahan suhu. Saat durasi penyinaran memiliki derajat yang tinggi, atap akan terbuka untuk penyinaran langsung dari matahari. Prediksi dari model tanpa transformasi *log-odds* pada Subbab E menunjukkan bahwa prediksi dapat bernilai lebih dari satu. Jika model prediksi ini diaplikasikan pada sistem rumah kaca dari penelitian Saadon et al. (2021), penyesuaian sudut atap tidak akan berjalan karena ketidaksesuaian nilai durasi penyinarannya. Seperti yang diketahui dari persamaan 8.2 dan persamaan 8.3, durasi penyinaran (σ_n) memiliki signifikansi yang tinggi terhadap sudut matahari (γ_s). Nilai prediksi yang melebihi satu akan membuat atap terus terbuka dan memperbesar laju evaporasi dari tanaman pangan (Choab et al., 2021). Hasilnya nutrisi akan teruapkan dan menurunkan performa tanaman pangan untuk berproduksi.

2. Implifikasi Manajerial

Perubahan iklim yang cukup ekstrem dan tingginya ketidakpastian (*uncertainty*) dari prakiraan cuaca (Jeantet et al., 2023) akan menjadi tantangan ke depannya untuk praktisi, petani, pemerhati, dan pemangku kebijakan untuk ketahanan pangan. Tidak seperti komoditas lainnya, tanaman pangan termasuk komoditas yang sensitif terhadap perubahan lingkungan. Beberapa komoditas pangan, seperti apel (Choab et al., 2021) cukup sensitif terhadap suhu. Tingginya suhu akan mempercepat reaksi kimia dari apel yang mempercepat tingkat kematangan buahnya. Dengan adanya perubahan iklim yang mengarah kepada cuaca bersuhu tinggi dan kering, kematangan buah nantinya akan lebih cepat. Cepatnya kematangan buah berdampak kepada penyesuaian kapasitas panen atau produksi yang juga harus bersifat cepat. Selain itu, keringnya cuaca akan mempercepat laju evaporasi dari tanaman pangan. Alhasil tanaman pangan yang matang akan dipanen dengan nutrisi yang rendah. Dari hal ini perlunya prediksi cuaca yang akurat dan tepat untuk proses pemanenan tanaman pangan yang lebih efektif dan efisien.

Kelembapan adalah salah satu variabel di prakiraan cuaca. Pentingnya prediksi yang akurat untuk variabel ini agar mengetahui waktu tepat untuk penyemprotan tanaman. Pada proses penanaman tanaman pangan, petani perlu mengetahui waktu saat kelembapan akan rendah. Rendahnya kelembapan mengartikan rendahnya kandungan air di udara sekitar dan akan mempercepat laju evaporasi dari tanaman (Liang et al., 2020). Pada prediksi di Kota Bandung, ada sekitar 5 periode dalam 5 tahun, daerah tersebut akan mengalami kelembapan yang cukup rendah. Dengan mengetahui secara akurat estimasi prediksi kelembapan akan rendah maka petani dapat menjadwalkan penyemprotan tanaman pangan lebih baik dan optimal. Prediksi kelembapan juga harus disesuaikan dengan kaidah matematika dan statistika. Hal ini berguna untuk proses pengambilan keputusan yang lebih mudah untuk diproses, dibandingkan data-data prediksi yang tidak memenuhi kaidah matematika.

Selain kelembapan, artikel ini juga membahas durasi penyinaran sebagai variabel dalam prakiraan cuaca. Prediksi durasi penyinaran juga sama pentingnya untuk para petani dalam menjadwalkan penyemprotan disaat durasi penyinaran mendekati 100%. Saat matahari memiliki durasi penyinaran yang tinggi, suhu dari lingkungan juga akan tinggi (Bartoszek & Matuszko, 2021). Tingginya suhu akan mempercepat evaporasi dari air dan nutrisi tanaman. Hal ini yang menyebabkan petani akan melakukan penyemprotan disaat durasi penyinaran yang tinggi. Selain itu, untuk tanaman pangan yang terkhusus dikembangkan di rumah kaca (*greenhouse*), durasi penyinaran perlu diketahui untuk mengatur kapasitas dan ventilasi dari rumah kaca. Tingginya durasi penyinaran akan memperkecil ventilasi untuk mengurangi laju evaporasi. Durasi penyinaran yang tinggi juga akan membuat petani untuk menambah kapasitas rumah kaca agar energi panas yang tercipta menyebar secara keseluruhan (tidak terfokus pada satu tempat). Sama seperti kelembapan, durasi penyinaran di kota Bandung memiliki probabilitas tinggi sebanyak 5 kali dalam 5 tahun.

Perubahan iklim dan cuaca yang tidak menentu mendorong beberapa peneliti dan akademisi untuk menciptakan teknologi dan

model prediksi cuaca untuk mengantisipasi ketahanan pangan (*food security*). Sensitivitas tanaman pangan terhadap cuaca dan lingkungan mendorong peneliti untuk memprediksikan cuaca dalam waktu berjangka pendek dan panjang. Selain untuk kegiatan operasional dalam penanaman tanaman pangan, seperti waktu penyiraman dan pengaturan ventilasi dari rumah kaca, prediksi cuaca dapat juga berguna untuk perencanaan kapasitas dan waktu pembibitan tanaman pangan. Prediksi deret waktu pada cuaca yang berkelembapan tinggi dan durasi penyinaran rendah menjadi waktu yang tepat untuk melakukan pembibitan tanaman karena laju evaporasi dari wilayah tersebut akan rendah. Juga mengetahui rata-rata kelembapan dan durasi penyinaran untuk tahun-tahun berikutnya akan berguna untuk perencanaan pemakaian sawah atau kebun terbuka atau rumah kaca untuk menanam tanaman pangan. Prediksi yang tidak memenuhi kaidah matematika dan statistika seperti model tanpa transformasi *log-odds* akan menyulitkan proses pengambilan keputusan kegiatan yang disebutkan.

G. Penutup

Bab ini telah menguraikan teknik transformasi data pada prakiraan cuaca dengan fungsi *log-odds*. Contoh pengolahan data kelembapan relatif dan durasi penyinaran, telah menunjukkan bahwa transformasi *log-odds* dapat membantu proses prediksi dalam mengikuti kaidah ataupun kriteria parameter iklim. Tanpa transformasi *log-odds*, data prediksi menunjukkan adanya nilai negatif dari kelembapan relatif dan nilai yang melebihi 100% dari durasi penyinaran. Padahal, dalam perhitungan data aktual, nilai negatif dari kelembapan relatif dan nilai durasi penyinaran yang lebih tinggi dari 100%, tidak mungkin terjadi. Hal ini dapat dicermati kembali dari pengertian masing-masing parameter. Namun, keberhasilan metode transformasi *log-odds* dalam "memandu" nilai prediksi agar sesuai dengan kaidah matematis parameter kelembapan relatif dan durasi penyinaran, tidak berarti bahwa metode tersebut berkontribusi pada peningkatan akurasi maupun resolusi data. Antara data prediksi dengan transformasi dan tanpa

transformasi, keduanya memiliki pola yang sama dengan deviasi yang relatif kecil.

Apabila dicermati, kemampuan "menangkap" pola musiman dari kedua parameter tersebut, mungkin ditentukan oleh model TSLM. Pada saat melakukan *training* dengan periode 2000–2021, terjadi *overfitting* antara data awal dengan hasil prediksi dengan transformasi. Namun, model TSLM gagal ketika dilakukan validasi untuk periode 2022. Kegagalan tersebut terlihat dari dua hal, yaitu dari pola yang terbentuk dan nilai hasil prediksi jangka panjang yang berada di luar kaidah. Oleh karena itu, terdapat peluang diperoleh hasil yang lebih optimal apabila digunakan perpaduan metode proyeksi dan transformasi yang tepat, melalui serangkaian proses pemilihan model.

Model dan prediksi iklim berbasis statistik, memang pada umumnya, tidak dapat sebaik model-model dinamis yang telah mempertimbangkan proses-proses fisis yang kompleks yang terjadi di alam, tetapi dalam beberapa kasus, model dan prediksi iklim berbasis statistik yang "penuh dengan penyederhanaan" tetap diperlukan. Analisis deret waktu untuk prediksi kelembapan relatif dan durasi penyinaran yang menjadi contoh dalam demonstrasi ini, menunjukkan banyak keterbatasan hasil karena terlalu banyak asumsi yang digunakan. Namun, penggunaan model-model prediksi nonlinear (dengan *time series logistic regression*, jaringan saraf tiruan (*artificial neural network*), mesin pendukung vektor (*support vector machine*), klasifikasi berbasis pohon (*classification-based tree*), dan lainnya) memiliki peluang untuk menghasilkan akurasi yang lebih tinggi. Selain itu, penggunaan variabel cuaca lainnya dengan tipe kejadian atau kelas yang lain akan lebih menantang karena kekompleksitasan proses fisisnya yang perlu dipertimbangkan dalam pemodelan iklim dan cuaca.

Referensi

- Aydinli, S., & Krochmann, J. (1987). Guide on daylighting of building interiors. Dalam *Data on daylight and solar radiation, Draft, CIE*.
- Badan Pusat Statistik. (2022a). *Kelembapan udara per bulan di Kota Bandung (persen), 2020-2022*. BPS. <https://bandungkota.bps.go.id/indicator/151/1249/1/kelembapan-udara-per-bulan-di-kota-bandung.html>
- Badan Pusat Statistik. (2022b). *Penyinaran matahari per bulan di Kota Bandung (Persen), 2020-2022*. BPS. <https://bandungkota.bps.go.id/indicator/151/1250/1/penyinaran-matahari-per-bulan-di-kota-bandung.html>
- Bartoszek, K., & Matuszko, D. (2021). The influence of atmospheric circulation over Central Europe on the long-term variability of sunshine duration and air temperature in Poland. *Atmospheric Research*, 251(December 2020), 105427. <https://doi.org/10.1016/j.atmosres.2020.105427>
- Black, K. (2023). *Business statistics: For contemporary decision making*. John Wiley & Sons.
- Chatfield, C., & Xing, H. (2019). *The analysis of time series: An introduction with R*. CRC Press.
- Choab, N., Allouhi, A., Maakoul, A. El, Kousksou, T., Saadeddine, S., & Jamil, A. (2021). Effect of greenhouse design parameters on the heating and cooling requirement of greenhouses in Moroccan climatic conditions. *IEEE Access*, 9, 2986–3003. <https://doi.org/10.1109/ACCESS.2020.3047851>
- Dobin, A. (2010). Data mining techniques for the life sciences. Dalam O. Carugo & F. Eisenhaber (Eds.), *Springer protocols - methods in molecular biology* 609 (Vol. 609, Issue January 2010). Humana Press. <https://doi.org/10.1007/978-1-60327-241-4>
- Hastie, T., Friedman, J., & Tibshirani, R. (2001). Additive models, trees, and related methods. Dalam *The elements of statistical learning: Data mining, inference, and prediction* (257–298). Springer. https://doi.org/10.1007/978-0-387-21606-5_9
- Heino, M., Kinnunen, P., Anderson, W., Ray, D. K., Puma, M. J., Varis, O., Siebert, S., & Kumm, M. (2023). Increased probability of hot and dry weather extremes during the growing season threatens global crop yields. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-29378-2>

- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). John Wiley & Sons. <https://doi.org/10.1002/9781118548387>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. Otexts. <https://otexts.com/fpp2/>
- Jeantet, A., Thirel, G., Lemaitre-Basset, T., & Tournebize, J. (2023). Uncertainty propagation in a modelling chain of climate change impact for a representative French drainage site. *Hydrological Sciences Journal*, 68(10), 1426–1442. <https://doi.org/10.1080/02626667.2023.2203322>
- Johansson, I. (2010). Metrological thinking needs the notions of parametric quantities, units and dimensions. *Metrologia*, 47(3), 219–230. <https://doi.org/10.1088/0026-1394/47/3/012>
- Korotcenkov, G. (2018). *Handbook of humidity measurement, Volume 1: Spectroscopic methods of humidity measurement*. CRC Press.
- Korotcenkov, G. (2020). *Handbook of Humidity Measurement, Volume 3: Sensing materials and technologies*. CRC Press.
- Liang, Y., Tabler, G. T., & Dridi, S. (2020). Sprinkler technology improves broiler production sustainability: From stress alleviation to water usage conservation: A mini review. *Frontiers in Veterinary Science*, 7(9), 1–8. <https://doi.org/10.3389/fvets.2020.544814>
- Maafi, A., & Adane, A. (1998). Analysis of the performances of the first-order two-state Markov model using solar radiation properties. *Renewable Energy*, 13(2), 175–193. [https://doi.org/10.1016/S0960-1481\(97\)00094-3](https://doi.org/10.1016/S0960-1481(97)00094-3)
- McClymont, H., Si, X., & Hu, W. (2023). Using weather factors and Google data to predict COVID-19 transmission in Melbourne, Australia: A time-series predictive model. *Heliyon*, 9(3), artikel e13782. <https://doi.org/10.1016/j.heliyon.2023.e13782>
- Nagaraj, R., & Kumar, L. S. (2023). Univariate deep learning models for prediction of daily average temperature and relative humidity: The case study of Chennai, India. *Journal of Earth System Science*, 132(3). <https://doi.org/10.1007/S-12040-023-02122-0>
- Nagarajan, G., & Minu, R. I. (2018). Wireless soil monitoring sensor for sprinkler irrigation automation system. *Wireless Personal Communications*, 98(2), 1835–1851. <https://doi.org/10.1007/S-11277-017-4948-y>

- Panesar, A. (2021). Machine learning algorithms. Dalam *Machine learning and AI for healthcare* (85–144). Apress. https://doi.org/10.1007/978-1-4842-6537-6_4
- Paolella, M. S. (2018). *Linear models and time-series analysis: Regression, ANOVA, ARMA and GARCH*. Wiley. <https://doi.org/10.1002/9781119432036>
- Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., De Baets, S., Dokumentov, A., ... Ziel, F. (2022). Forecasting: Theory and practice. *International Journal of Forecasting*, 38(3), 705–871. <https://doi.org/10.1016/j.ijforecast.2021.11.001>
- Rahim, R., Baharuddin, & Mulyadi, R. (2004). Classification of daylight and radiation data into three sky conditions by cloud ratio and sunshine duration. *Energy and Buildings*, 36(7), 660–666. <https://doi.org/10.1016/j.enbuild.2004.01.012>
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Rumble, J. (Ed.). (2017). *CRC handbook of chemistry and physics* (98th ed.). CRC Press.
- Saadon, T., Lazarovitch, N., Jerszurki, D., & Tas, E. (2021). Predicting net radiation in naturally ventilated greenhouses based on outside global solar radiation for reference evapotranspiration estimation. *Agricultural Water Management*, 257(August), 107102. <https://doi.org/10.1016/j.agwat.2021.107102>
- Sammut, C., & Webb, G. I. (Ed.). (2010). *Encyclopedia of machine learning*. Springer. <https://doi.org/10.1007/978-0-387-30164-8>
- Santos, C. M. dos, Escobedo, J. F., de Souza, A., da Silva, M. B. P., & Aristone, F. (2021). Prediction of solar direct beam transmittance derived from global irradiation and sunshine duration using anfis. *International Journal of Hydrogen Energy*, 46(55), 27905–27921. <https://doi.org/10.1016/j.ijhydene.2021.06.044>

- Sridhara, S., Manoj, K. N., Gopakkali, P., Kashyap, G. R., Das, B., Singh, K. K., & Srivastava, A. K. (2023). Evaluation of machine learning approaches for prediction of pigeon pea yield based on weather parameters in India. *International Journal of Biometeorology*, 67(1), 165–180. <https://doi.org/10.1007/s00484-022-02396-x>
- Tabari, H., & Willems, P. (2023). Global risk assessment of compound hot-dry events in the context of future climate change and socioeconomic factors. *Npj Climate and Atmospheric Science*, 6(1). <https://doi.org/10.1038/s41612-023-00401-7>