



Perbandingan Metode *Supervised Machine Learning* untuk Prediksi Kelembapan Tanah di Jakarta

Firda Azzahrotunnisa¹, Mohammad Ilham Mahfud², Nurin Salsabila Ishak³, Robert
Kurniawan⁴

^{1,3}Prodi Statistika, Politeknik Statistika STIS, Jakarta, Indonesia

⁴Prodi Komputasi Statistik, Politeknik Statistika STIS, Jakarta, Indonesia

¹212112061@stis.ac.id

³212112274@stis.ac.id

⁴robertk@stis.ac.id

²Prodi Statistika, Politeknik Statistika STIS, Jakarta, Indonesia

Corresponding author email: ²212112192@stis.ac.id

Abstract: Water is the most important resource for the survival of living things. Along with the increasing population and population activities, the availability of clean water is quite worrying. This study aims to predict soil moisture using Random Forest, Neural Network, and SVM methods with ERA5-Land data. The results showed that the random forest method produced the highest R-Squared of 70.18% with a MAPE of 0.0302 and a RMSE of 0.0158 so that the model could predict soil moisture quite well. The Random Forest algorithm has proven to be the best model for predicting soil moisture. This can subsequently aid in groundwater management and inform policies regarding groundwater.

Keywords: Soil Moisture, Neural Network, Random Forest, SVM

Abstrak: Air merupakan sumber daya terpenting untuk keberlangsungan makhluk hidup. Seiring dengan meningkatnya populasi dan aktivitas penduduk, ketersediaan air bersih cukup mengkhawatirkan. Penelitian ini bertujuan memprediksi kelembapan tanah yang dilakukan menggunakan metode *Random Forest*, *Neural Network*, dan SVM dengan data *ERA5-Land*. Hasil menunjukkan metode *Random Forest* menghasilkan *R-Squared* tertinggi 70,18% dengan MAPE 0,0302 dan RMSE 0,0158 sehingga model dapat memprediksi kelembapan tanah dengan cukup baik. Algoritma *Random Forest* menjadi model terbaik untuk melakukan prediksi kelembapan tanah. Hal ini nantinya dapat membantu dalam pengelolaan air tanah dan menindaklanjuti kebijakan mengenai air tanah.

Kata kunci: Kelembapan Tanah, Neural Network, Random Forest, SVM

I. PENDAHULUAN

Air merupakan sumber daya terpenting bagi kelangsungan hidup manusia. Air tanah, sebagai salah satu sumber air, memegang peran yang sangat penting dalam hal ini. Air tanah dihubungkan dengan kelembapan tanah untuk mengukur kadar air tanah di suatu daerah. Kelembapan tanah adalah sejumlah kadar air yang terdapat diantara partikel – partikel yang terkandung di dalam matriks tanah [1]. Matriks tanah merupakan kumpulan tanah yang terdiri dari partikel-partikel padat serta rongga-rongga yang diisi oleh air dan udara.

Seiring dengan meningkatnya aktivitas manusia dan terjadinya perubahan iklim dunia, kualitas dan ketersediaan air tanah cukup mengkhawatirkan, khususnya di daerah padat penduduk. Menurut Badan Pusat Statistik (BPS) [2], kepadatan penduduk Indonesia pada tahun 2021 mencapai 142 jiwa/km² dengan provinsi tertinggi adalah Provinsi DKI Jakarta, yakni 15.978 jiwa/km². Hal ini tak terlepas dari DKI Jakarta yang menjadi kota metropolitan dengan bisnis dan industri terbesar di Indonesia. Dengan demikian yang memicu datangnya perantau untuk mencari pekerjaan dan peluang usaha yang dapat berkontribusi pada kepadatan penduduk. Peningkatan kegiatan manusia ini berdampak pada pembangunan yang masif dan berkurangnya lahan resapan alami, yang berakibat pada menurunnya kualitas dan kelembapan tanah. Sementara itu, permintaan akan kebutuhan air tanah akan cenderung



meningkat seiring bertambahnya populasi penduduk, didukung dengan adanya perubahan iklim akan menambah beban akan ketersediaan air tanah [3].

Penurunan kualitas dan kelembapan tanah dapat berpengaruh pada ketahanan kekeringan dan pengendalian banjir [4]. Selain itu, penggunaan air tanah yang berlebihan dapat memberikan efek buruk terhadap lingkungan, salah satunya penurunan muka air tanah. Eksploitasi air tanah secara berlebihan dapat meninggalkan kekosongan dalam ruang yang sebelumnya diisi oleh air tanah, sehingga penyangga lapisan akan hilang dan permukaan tanah akan turun secara berangsur.

Teknik *machine learning* seringkali digunakan untuk pengolahan pada big data yang luas dan beragam hasil penginderaan jarak jauh, dengan performa yang sangat baik [5]. Penelitian oleh Gill, M.K., *et al.* [6] menunjukkan bahwa prediksi kelembapan tanah dengan *Support Vector Machine* (SVM) lebih baik dibanding *Artificial Neural Network* (ANN). Sedangkan pada penelitian yang dilakukan oleh Sagarika Paul dan Satwinder Singh [7], SVM menjadi metode terbaik dalam prediksi kelembapan tanah dibandingkan teknik *machine learning* lain. Prakasha *et al.* [8] menemukan bahwa *Multiple Linear Regression* (MLR) lebih baik dibandingkan SVM dan *Recurrent Neural Network* (RNN). Penelitian oleh Carranza, C. *et al.* [9] menggunakan metode *Random Forest* (RF) dalam mengestimasi kelembapan tanah tetapi hasilnya kurang akurat yang disebabkan karena kurangnya pengambilan sampel pada lokasi penelitian dan representasi yang tidak lengkap mengenai proses bagian bawah permukaan. Sehingga, dalam penelitian ini digunakan beberapa model *machine learning* yang menjadi pertimbangan metode untuk memprediksi kelembapan tanah, seperti RF, SVM, dan NN.

Pengukuran kelembapan tanah ini menjadi sangat penting untuk berbagai penerapan hidrometeorologi diantaranya seperti alokasi sumber daya air [10], manajemen irigasi [11], dan prediksi bencana kekeringan dan banjir [12], [13]. Oleh karena itu, diperlukan prediksi kelembapan tanah secara akurat agar dapat menjadi pertimbangan kebijakan pemerintah dalam upaya menjaga ketersediaan dan kualitas air tanah serta pengelolaan sumber daya air yang efektif. Penelitian ini bertujuan untuk memprediksi tingkat kelembapan tanah lapisan 1 (0-7 cm) per jam di Sta. Met. Kelas III Kemayoran - Jakarta Pusat. Diharapkan dengan adanya analisis ini dapat dibentuk kebijakan yang mampu membantu penyerapan dan pengelolaan air di daerah sekitar stasiun.

II. METODE PENELITIAN

2.1 Kelembapan Tanah

Kelembapan tanah atau *soil moisture* (SM) mengacu pada jumlah air yang terkandung dalam *unsaturated soil zone* atau *vadose zone* [14]. SM merupakan variabel hidrologi yang mengatur sirkulasi air dan energi antara atmosfer dan tanah. Kelembapan tanah menjadi komponen penyimpanan curah hujan terbesar dan anomali radiasi di darat [15]. Air yang terkandung dalam tanah akan berpindah ke atmosfer melalui proses evapotranspirasi dari daratan yang meliputi transpirasi tanaman dan penguapan tanah gundul [16]. Sebesar 60% dari curah hujan yang turun ke daratan akan kembali ke atmosfer melalui evapotranspirasi [16]. Selain itu, evapotranspirasi juga melibatkan perpindahan energi sehingga disebut juga sebagai fluks energi. Lebih dari 50% total energi matahari yang diserap permukaan daratan akan digunakan untuk evapotranspirasi. Pertukaran energi ini terbagi menjadi dua bentuk yaitu *flux heat laten* dan *flux heat sensible* [17]. *Flux heat laten* merupakan perpindahan energi dikarenakan terjadinya proses penguapan. *Flux heat sensible* adalah perpindahan energi yang terjadi karena perbedaan suhu antara atmosfer dan permukaan. Semakin besar perbedaan suhu maka aliran energinya juga akan semakin besar. Energi akan mengalir dari atmosfer ke daratan jika suhu atmosfer lebih hangat begitupun sebaliknya.



SM memiliki hubungan yang kuat dengan perubahan iklim, variasi suhu, curah hujan dan faktor iklim lainnya. Penelitian oleh Hirschi *et al.*, (2011) menunjukkan bahwa korelasi antara kelembapan tanah dan suhu ekstrem sangat kuat [18]. Peningkatan suhu pada kulit permukaan akan menurunkan kelembapan tanah selama masa kekeringan [19]. Selain itu, kelembapan tanah juga sangat bergantung pada penggunaan lahan atau penutupan lahan. Perubahan dalam penggunaan lahan/tutupan lahan akan berpengaruh terhadap iklim global termasuk curah hujan [20].

Pengetahuan mengenai kelembapan tanah sangat penting untuk berbagai aplikasi di bidang meteorologi, klimatologi, dan hidrologi. Prakiraan mengenai kelembapan tanah juga membantu di bidang pertanian, diantaranya sebagai penjadwalan irigasi, pemantauan kualitas air, bahkan prakiraan hasil panen. Informasi kelembapan tanah ini juga dapat dimanfaatkan untuk studi cuaca dan perubahan iklim, pemantauan dan prakiraan cuaca ekstrem.

Pengukuran kelembapan tanah telah dilakukan secara langsung (*in situ*) dengan teknik gravimetri, nuklir, elektromagnetik dan higrometer. Ada berbagai model berbasis fisik untuk mengestimasi kelembapan tanah, tetapi penggunaan model berbasis fisik ini mengalami kesulitan dalam memperkirakan parameter fisik yang diperlukan. Teknik penginderaan jarak jauh juga telah digunakan dalam pengukuran kelembapan tanah, tetapi dalam pelaksanaannya memerlukan biaya dan upaya yang besar. Kesulitan dengan pendekatan secara fisik memaksa peneliti mencari alat peramalan berbasis data menggunakan model statistik dan pemanfaatan *machine learning*.

2.2 Data

Data yang digunakan dalam penelitian ini berasal dari kumpulan data *ERA 5-Land Climate Reanalysis* yang merupakan data meteorologi oleh *Copernicus Climate Service*. *ERA 5-Land* diproduksi dengan memutar ulang komponen lahan dari analisis ulang iklim *ECMWF ERA 5*. Data ini mencakup global per jam dengan resolusi horizontal 9 km. Lokus yang digunakan yakni Sta. Met. Kelas III Kemayoran - Jakarta Pusat yang terletak pada 106,84 BT dan 6,17 LS dengan rentang waktu dari 1 Mei 2024 hingga 24 Mei 2024.

Penelitian ini menggunakan delapan variabel, di antaranya 5 variabel berupa data meteorologi seperti suhu 2m, suhu titik embun 2 meter, total curah hujan, komponen angin 10 m ke arah timur dan komponen angin 10m ke arah utara, sedangkan sisanya terkait dengan karakteristik tanah seperti penguapan dari tanah kosong, flux panas sensibel permukaan, dan kelembapan tanah.

Kelembapan tanah pada data *ERA 5-Land* terlokalisasi dalam 4 lapisan yaitu lapisan 1 kedalaman 0-7 cm, lapisan 2 kedalaman 7-28 cm, lapisan 3 kedalaman 28-100 cm, dan lapisan 4 kedalaman 100-289 cm. Lapisan 1 dari kelembapan tanah dipilih dalam penelitian ini karena memiliki kedalaman yang paling dekat dengan permukaan tanah, yang membuatnya sensitif terhadap perubahan kondisi meteorologi dan faktor lingkungan. Keterangan mengenai variabel yang digunakan dalam penelitian tertera pada Tabel 1.

Tabel 1. Keterangan Variabel

Variabel	Satuan	Keterangan
Suhu 2m (t2m)	K	Suhu udara pada ketinggian 2 meter di atas permukaan tanah, laut atau perairan pedalaman.
Suhu titik embun 2 meter (d2m)	K	Suhu udara yang didinginkan agar terjadi kejenuhan pada ketinggian 2 meter di atas permukaan bumi.
Total Curah Hujan (tp)	m	Akumulasi air cair dan beku, termasuk hujan dan salju, yang jatuh ke permukaan bumi.



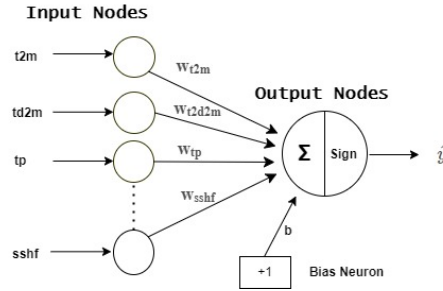
Variabel	Satuan	Keterangan
Komponen Angin 10 m ke arah Timur (u_{10})	$m s^{-1}$	Kecepatan horizontal udara yang bergerak ke arah timur, pada ketinggian 10 meter di atas permukaan bumi dalam meter per detik.
Komponen angin 10 m ke arah utara (v_{10})	$m s^{-1}$	Kecepatan horizontal udara yang bergerak ke arah utara, pada ketinggian 10 meter di atas permukaan bumi, dalam meter detik
Penguapan dari tanah kosong (evaps)	m air setara	Besarnya penguapan dari tanah gundul.
Fluks panas sensibel permukaan (sshf)	$J m^{-2}$	Perpindahan panas antara permukaan bumi dan atmosfer melalui efek gerakan turbulen udara (tetapi tidak termasuk perpindahan panas akibat kondensasi atau penguapan).
Kelembapan tanah (swv_{11})	$m^3 m^{-3}$	Volume air pada lapisan tanah 1 (0 - 7 cm).

2.3 Random Forest (RF)

Algoritma RF pertama kali diusulkan oleh L. Breiman pada tahun 2001 yang secara sukses mampu menyelesaikan permasalahan klasifikasi dan regresi [21]. RF adalah algoritma *ensemble learning* yang menggabungkan konsep *decision tree* dan *bagging* [22]. *Decision Tree* (DT) adalah *supervised learning* non-parametrik digunakan untuk klasifikasi ataupun regresi. DT ini bekerja dengan mempartisi ruang variabel menggunakan seperangkat aturan hierarki sehingga data akan dikelompokkan secara rekursif. Sekumpulan kovariat digunakan untuk memisahkan nilai variabel secara rekursif sehingga menghasilkan beberapa node induk dan anak yang menyerupai struktur pohon. Bagging (*bootstrap aggregation*) adalah teknik untuk menghasilkan *data training* dengan melakukan *resampling with replacement* pada *training set*. Hal ini berarti beberapa data mungkin akan digunakan lebih dari satu kali dalam training sementara data lainnya tidak pernah digunakan. Untuk setiap sampel bootstrap, DT akan menghasilkan beberapa node induk dan anak hingga kriteria penghentiannya tercapai. RF memiliki beberapa keunggulan diantaranya adalah kemampuan untuk menemukan hubungan non linier dalam data berdimensi tinggi, tahan terhadap overfitting, dan relatif tahan terhadap noise dalam prediktor [23].

2.4 Neural Network (NN)

NN merupakan bagian dari *Artificial Intelligence* (AI) yang dirancang untuk dapat memproses informasi dengan meniru cara kerja sel saraf pada otak manusia [25]. NN sederhana pertama kali diperkenalkan oleh McCulloch dan Pitts pada tahun 1943. NN berfungsi sebagai alat eksplorasi seperti estimasi fungsi non-linier, data sorting, mendeteksi pola tertentu, optimasi, clustering dan simulation [26]. Pada penerapan prediksi, NN dianggap sebagai *a nonlinear black box model* (Input-Output) [27]. Desain dari NN terdiri atas layer: *input-layer*, *hidden-layer*, *output-layer*, bobot dan *bias* serta *summation node*. NN merupakan struktur unit individu yang saling berhubungan yang disebut sebagai neuron. Setiap neuron akan menerima input signal, kemudian memprosesnya dan mengirimkan signal output.



Gambar 1. Komponen *Neural Network*

Input akan masuk melalui input-layer kemudian akan dihubungkan ke setiap neuron pada hidden-layer 1 seterusnya sampai pada output-layer. Bobot dan bias b_j dijumlahkan net input I_j pada persamaan (1). Kemudian akan diteruskan ke hidden-layer dengan fungsi pada persamaan (2) untuk menghasilkan output pada persamaan (3) [27].

$$I_j = \sum_{i=1}^d w_{ij} x_i + b_j \quad (1)$$

$$\text{sign}(I) = \frac{1-e^{-2I}}{1+e^{-2I}} \quad (2)$$

$$\hat{y} = \text{sign}(I_j) \quad (3)$$

Fungsi sign pada (2) akan bertindak sebagai fungsi aktivasi. Fungsi ini digunakan untuk mensimulasikan berbagai jenis model seperti *least-square regression*, *svm*, atau *logistic regression classifier*.

2.5 Support Vector Machine Algorithm (SVM)

SVM ditemukan pertama kali oleh Vapnik pada tahun 1979. Secara umum, SVM dibagi menjadi dua bentuk yaitu *Super Vector Classification (SVC)* dan *Support Vector Regression (SVR)* [9]. SVM yang digunakan pada penelitian ini berupa SVR. Pada SVR, fungsi keputusan linear didefinisikan sebagai $f(x) = \langle \mathbf{w}, \mathbf{x} \rangle + b$ dimana $\langle \mathbf{w}, \mathbf{x} \rangle$ adalah dot product antara vektor pembobot \mathbf{w} dan vektor input \mathbf{x} ; dan b adalah bias. Pengamatan tersebut disebut fungsi regulasi seperti sebagai berikut:

$$\begin{aligned} &\text{Meminimumkan } \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^L (\xi_i + \xi_i^*) \\ &\text{pada } \begin{cases} y_i - \sum_{j=1}^K \sum_{i=1}^L w_j x_{ji} - b \leq \varepsilon + \xi_i, \\ \sum_{j=1}^K \sum_{i=1}^L w_j x_{ji} + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, \end{cases} \quad (4) \end{aligned}$$

dimana ε adalah fungsi Vapniks insensitive loss; C adalah parameter capacity; ξ_i dan ξ_i^* adalah variabel slack. Hal ini berarti SVR mengabaikan error yang ditimbulkan oleh data yang dibatasi dalam margin ε dan mempertimbangkan sisanya untuk menemukan *hyperplane optimal* dengan bantuan variabel slack ξ_i . Sehingga fungsi yang digunakan dapat ditulis sebagai berikut

$$f(x) = \sum_{i=1}^N (\alpha_i^* - \alpha_i) k(x, x_i) + b \quad (5)$$

$$\sum_{i=1}^L (\alpha_i^* - \alpha_i) = 0 \text{ dan } 0 \leq \alpha_i, \alpha_i^* \leq C \quad (6)$$



Dimana $i=1,\dots,L$ adalah ukuran sampel dan α_i adalah lagrange multipliers. Pada persamaan (5), $k(x, x_i)$ adalah fungsi kernel. Dalam penelitian ini, fungsi kernel yang digunakan adalah kernel radial seperti yang disarankan oleh Scholkopf et al. (1997) karena memiliki performa terbaik dibandingkan kernel lainnya [31]. Formula kernel radial dapat ditulis sebagai berikut

$$k(x, x_i) = e^{-\gamma \|x-x_i\|^2} \quad (7)$$

Dimana γ adalah parameter kernel. Parameter lainnya ditentukan melalui prosedur trial and error yang bertujuan untuk mengevaluasi model.

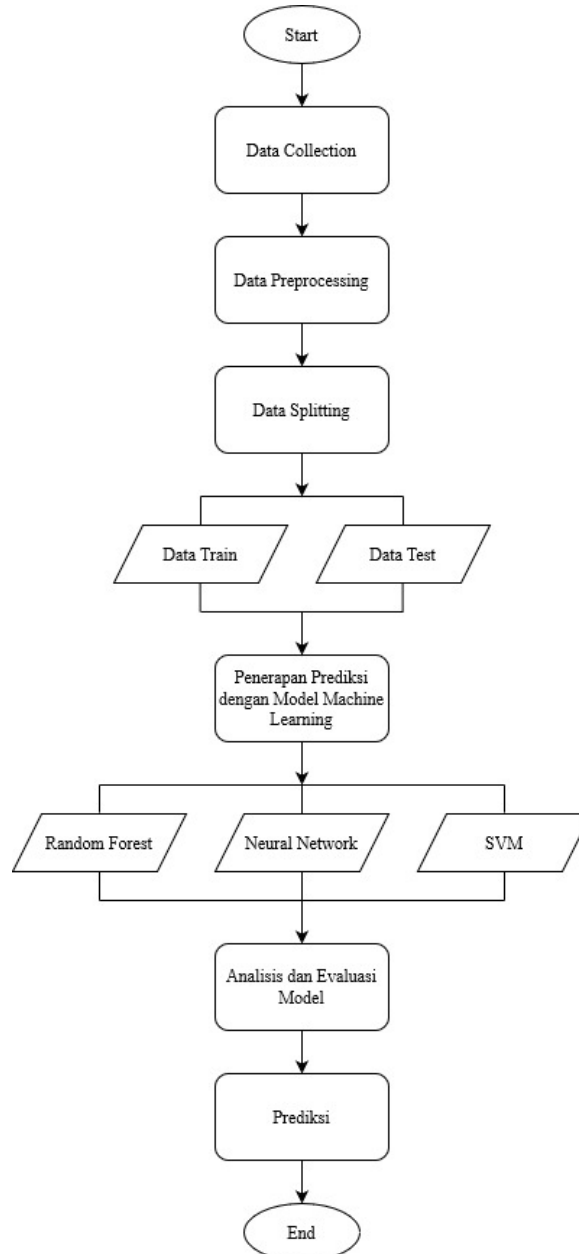
2.6 Evaluasi Model

Tiga ukuran evaluasi dipilih untuk menunjukkan kinerja model yang berbeda.

$$R \text{ Squared } (R^2) : 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (8)$$

$$\text{Root Mean Squared Error (RMSE)} : \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (9)$$

Pada persamaan (8) dan (9), \hat{y} adalah nilai prediksi, y adalah nilai sebenarnya, dan \bar{y} adalah nilai rata-rata.



Gambar 2. Tahapan Analisis

Berdasarkan Gambar 2, tahapan analisis yang dilakukan adalah sebagai berikut.

1. Melakukan pengumpulan data (*data collection*);
2. Melakukan *preprocessing data* untuk melihat *missing value*;
3. Melakukan *data splitting* menjadi *train data* dan *test data* dengan proporsi 80:20;
4. Membentuk model dari *train data* dengan menerapkan algoritma *machine learning* berupa RF, NN, dan SVM;
5. Melakukan pengujian pada model dengan *test data*;
6. Melakukan evaluasi model;



7. Model terbaik yang dipilih berdasarkan nilai R-Square dan RMSE selanjutnya akan dilakukan untuk membuat prediksi.

III. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari kumpulan data *ERA 5-Land Climate Reanalysis*. Data didapatkan dalam bentuk formal file *.nc* sehingga diperlukan pengolahan terlebih dahulu sebelum dilakukan pengolahan sebagai dataset analisis. Data yang diambil berupa data kelembapan tanah beserta variabel yang memengaruhinya.

3.2 Preprocessing Data

Pada tahapan ini, dilakukan pemrosesan pada data sebelum dilakukan pemodelan. Pemrosesan data awal dilakukan dengan melakukan ekstraksi dataset, yaitu merubah format file *.nc* ke dalam bentuk excel dengan menggunakan bantuan *Google Collab*. Selanjutnya, dilakukan *data cleaning* untuk mengecek apakah data memiliki nilai yang tidak valid atau *missing value*. Hasil menunjukkan bahwa tidak terdapat satupun nilai *missing*, sehingga 576 observasi digunakan secara keseluruhan dalam analisis. Kemudian dilakukan penyesuaian pada tipe data pada variabel untuk mempermudah kegiatan analisis.

3.3 Statistika Deskriptif

Karakteristik dan gambaran umum disajikan dalam Tabel 2, yang menunjukkan ringkasan statistik untuk setiap variabel yang digunakan dalam penelitian ini.

Tabel 2. Summary Statistics

Variabel	Rerata	Min	Max	St. Deviasi
u10	-0,0533	-4,4999	1,5442	1,0609
v10	0,1032	-2,7049	1,7637	1,1331
d2m	297,4	295,5	298,9	0,5821
t2m	301,0	297,8	305,0	1,9664
evabs	-2,083e-03	-3,330e-03	-9,740e-07	0,0009
sshf	-2.650.991	-4.545.108	24.063	1.119.860
tp	1,257e-03	0,000e+00	2,232e-02	0,0024

Hasil pada Tabel 2, menunjukkan bahwa pada bulan Mei angin pada ketinggian 10 meter cenderung bertiup ke arah barat (u10 rata-rata -1.0533) dan utara (v10 rata-rata 0.1032) dengan keragaman sebesar 1,06099 dan 1,1331. Temperatur titik embun pada 2 meter (d2m) memiliki nilai rata-rata 297.4 K, menunjukkan kelembapan udara yang tinggi, sementara temperatur udara pada 2 meter (t2m) memiliki nilai rata-rata 301.0 K, menunjukkan kondisi yang hangat. Evapotranspirasi (evabs) sangat rendah dengan rata-rata -2.083e-03. Fluks panas sensibel dari permukaan (sshf) menunjukkan perpindahan panas dari antara permukaan bumi dan atmosfer dengan rata-rata -2,650,991 W/m². Total presipitasi (tp) rata-rata 1.257e-03 m menunjukkan curah hujan yang rendah.

3.4 Pembagian Data

Data dibagi menjadi *train data* dan *test data*. *Train data* digunakan untuk pembelajaran pemodelan *machine learning*, sedangkan *test data* digunakan untuk menguji hasil dari pemodelan yang dihasilkan *machine learning*. Data dibagi dengan rasio 80% untuk *train data* dan 20% untuk *test data*.



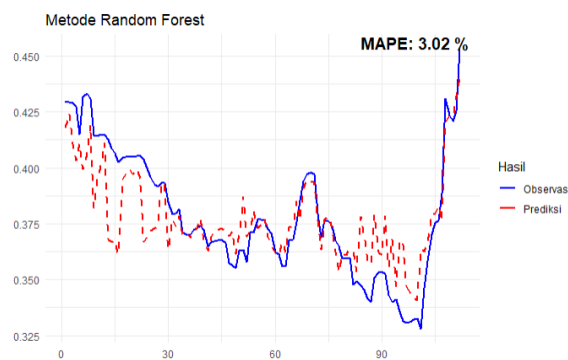
3.5 Evaluasi Model

Perbandingan algoritma *supervised machine learning* untuk memprediksi kelembapan tanah pada Tabel 3 menunjukkan bahwa algoritma *Random Forest* memiliki *R-squared* sebesar 70,18% dengan RMSE 0,0158. Pada algoritma *Neural Network*, didapatkan *R-squared* sebesar 63,97% dengan RMSE 0,0171. Adapun hasil prediksi dengan SVM memiliki *R-squared* sebesar 52,21% dengan RMSE 0,0197.

Tabel 3. Perbandingan Metode Supervised Machine Learning

No	Metode	R-Squared	RMSE
1	Random Forest	0.7018	0,0158
2	Neural Network	0.6397	0.0171
3	SVM	0.5221	0.0197

Metode *Random Forest* melakukan prediksi yang paling baik dibandingkan dengan metode *Neural Network* dan SVM dilihat dari performanya. Tingkat kesalahan yang dihasilkan dari prediksi metode *Random Forest* paling kecil yakni hanya sebesar 0,0158 dengan *R-squared* tertinggi yakni 70,18%. Hal ini sejalan dengan penelitian yang dilakukan Carranza *et al.*, (2021) yang memprediksi kelembapan tanah menggunakan *Random Forest*, menyatakan bahwa model RF memiliki akurasi yang lebih tinggi untuk memprediksi kelembapan tanah saat menggunakan data yang sudah ada dan akurasi yang hampir sama untuk memprediksi kondisi di luar data yang ada [9]. Penelitian lainnya oleh Jia *et al.* (2020) mengungkapkan bahwa model RF dapat memberikan estimasi kelembapan tanah yang baik bahkan dalam kondisi dimana informasi tentang jenis tanah tidak lengkap atau tidak tersedia [28]. Algoritma *Random Forest* mampu menangkap hubungan kelembapan tanah dengan efektif [29]. Efektivitas algoritma RF dapat dievaluasi dengan membandingkan hasil prediksinya dengan data observasi yang diamati.



Gambar. 3 Grafik Perbandingan Nilai Hasil Prediksi dan Observasi

Gambar 4 menunjukkan hasil prediksi berdasarkan model RF yang dihasilkan, dengan tingkat kesalahan yang dihasilkan untuk grafik prediksi sedikit buruk di awal dan mulai membaik di akhir. Hasil prediksi menunjukkan nilai kelembapan tanah di Stasiun Kemayoran, Jakarta Pusat. Tingkat kelembapan tanah sempat turun di bulan Mei 2024. Hal ini tak terlepas karena adanya fenomena el nino yang melanda Indonesia [30]. Namun, kelembapan tanah mulai merangkak naik pada akhir bulan Mei 2024.



IV. KESIMPULAN

Prediksi kelembapan tanah dengan metode *Random Forest* dapat menghasilkan prediksi yang cukup baik dengan *R-Squared* 70,18%. Penelitian ini dimudahkan dengan akses data yang mudah dan tersedia secara *real-time*, walaupun ada beberapa data yg kurang valid akibat tertutup awan dan asap polusi.

UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam penyelesaian penelitian ini. Secara khusus, kami berterima kasih kepada:

1. Bapak Robert Kurniawan, SST, M.Si yang telah memberikan bimbingan, arahan, dan dukungan yang tiada henti selama proses penelitian ini berlangsung.
2. Kelompok 1 dan teman-teman kelas 3SK2 yang telah memberikan saran, kritik, dan dukungan moral yang sangat berarti.
3. Keluarga dan sahabat yang telah memberikan dukungan moral dan motivasi selama masa penelitian ini.

Tanpa kontribusi dan dukungan dari semua pihak tersebut, penelitian ini tidak akan dapat diselesaikan dengan baik. Terima kasih disampaikan kepada Tim SENADA yang telah meluangkan waktu untuk mereview paper ini. Semoga hasil penelitian ini dapat memberikan manfaat dan kontribusi positif bagi ilmu pengetahuan dan masyarakat.

REFERENSI

1. Stasiun Klimatologi Sumatera Selatan, “Soil Moisture.” Accessed: May 27, 2023. [Online]. Available: <https://staklim-sumsel.bmkg.go.id/soil-moisture/>
2. BPS, “Kepadatan Penduduk menurut Provinsi (jiwa/km²), 2021,” 2021. [Online]. Available: <https://www.bps.go.id/id/statistics-table/2/MTQxIzI=/kepadatan-penduduk-menurut-provinsi--jiwa-km2-.html>
3. W. Y. Wu *et al.*, “Divergent effects of climate change on future groundwater availability in key mid-latitude aquifers,” *Nat. Commun.*, vol. 11, no. 1, pp. 1–9, 2020, doi: 10.1038/s41467-020-17581-y.
4. Y. Cai, W. Zheng, X. Zhang, L. Zhangzhong, and X. Xue, “(Paper 8 Tafur),” *PLoS One*, vol. 14, no. 4, pp. 1–19, 2019.
5. T. T. Nguyen *et al.*, “A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm,” *Sci. Total Environ.*, vol. 833, p. 155066, Aug. 2022, doi: 10.1016/j.scitotenv.2022.155066.
6. M. K. Gill, T. Asefa, M. W. Kemblowski, and M. McKee, “Soil moisture prediction using support vector machines,” *J. Am. Water Resour. Assoc.*, vol. 42, no. 4, pp. 1033–1046, 2006, doi: 10.1111/j.1752-1688.2006.tb04512.x.
7. S. Paul and S. Singh, “Soil Moisture Prediction Using Machine Learning Techniques,” *ACM Int. Conf. Proceeding Ser.*, pp. 1–7, 2020, doi: 10.1145/3440840.3440854.
8. S. Prakash, A. Sharma, and S. S. Sahu, “Soil Moisture Prediction Using Machine Learning,” *Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2018*, pp. 1–6, 2018, doi: 10.1109/ICICCT.2018.8473260.
9. C. Carranza, C. Nolet, M. Pezij, and M. van der Ploeg, “Root zone soil moisture estimation with Random Forest,” *J. Hydrol.*, vol. 593, no. November 2020, p. 125840, 2021, doi: 10.1016/j.jhydrol.2020.125840.
10. B. Grové, “Improved Water Allocation under Limited Water Supplies Using Integrated Soil-Moisture Balance Calculations and Nonlinear Programming,” *Water Resour. Manag.*, vol. 33, no. 1, pp. 423–437, 2019, doi: 10.1007/s11269-018-2110-6.
11. Y. Ouyang, G. Feng, T. D. Leininger, J. Read, and J. N. Jenkins, “Pond and Irrigation Model (PIM): a Tool for Simultaneously Evaluating Pond Water Availability and Crop Irrigation Demand,” *Water Resour. Manag.*, vol. 32, no. 9, pp. 2969–2983, 2018, doi: 10.1007/s11269-018-1967-8.



12. N. Chen *et al.*, “Drought propagation in Northern China Plain: A comparative analysis of GLDAS and MERRA-2 datasets,” *J. Hydrol.*, vol. 588, no. April, p. 125026, 2020, doi: 10.1016/j.jhydrol.2020.125026.
13. W. Pei, Q. Fu, D. Liu, T. Li, K. Cheng, and S. Cui, “A Novel Method for Agricultural Drought Risk Assessment,” *Water Resour. Manag.*, vol. 33, no. 6, pp. 2033–2047, 2019, doi: 10.1007/s11269-019-02225-8.
14. H. Vereecken, J. A. Huisman, H. Bogena, J. Vanderborght, J. A. Vrugt, and J. W. Hopmans, “On the value of soil moisture measurements in vadose zone hydrology: A review,” *Water Resour. Res.*, vol. 46, no. 4, pp. 1–21, 2008, doi: 10.1029/2008WR006829.
15. J. Liu and Z. Pu, “Does Soil Moisture Have an Influence on Near-Surface Temperature?,” *J. Geophys. Res. Atmos.*, vol. 124, no. 12, pp. 6444–6466, Jun. 2019, doi: 10.1029/2018JD029750.
16. S. I. Seneviratne *et al.*, “Investigating soil moisture-climate interactions in a changing climate: A review,” *Earth-Science Rev.*, vol. 99, no. 3–4, pp. 125–161, 2010, doi: 10.1016/j.earscirev.2010.02.004.
17. M. Li and Z. Ma, “Sensible and Latent Heat Flux Variability and Response to Dry–Wet Soil Moisture Zones Across China,” *Boundary-Layer Meteorol.*, vol. 154, no. 1, pp. 157–170, 2015, doi: 10.1007/s10546-014-9963-x.
18. M. Hirschi *et al.*, “Observational evidence for soil-moisture impact on hot extremes in southeastern Europe,” *Nat. Geosci.*, vol. 4, no. 1, pp. 17–21, 2011, doi: 10.1038/ngeo1032.
19. V. Lakshmi, T. J. Jackson, and D. Zehrhuhs, “Soil moisture-temperature relationships: Results from two field experiments,” *Hydrol. Process.*, vol. 17, no. 15, pp. 3041–3057, 2003, doi: 10.1002/hyp.1275.
20. R. A. Pielke *et al.*, “An overview of regional land-use and land-cover impacts on rainfall,” *Tellus, Ser. B Chem. Phys. Meteorol.*, vol. 59, no. 3, pp. 587–601, 2007, doi: 10.1111/j.1600-0889.2007.00251.x.
21. G. Biau and E. Scornet, “A random forest guided tour,” *TEST*, vol. 25, no. 2, pp. 197–227, Jun. 2016, doi: 10.1007/s11749-016-0481-7.
22. Leo Breiman, “Random Forest,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
23. H. Adab, R. Morbidelli, C. Saltalippi, M. Moradian, and G. A. F. Ghalhari, “Machine learning to estimate surface soil moisture from remote sensing data,” *Water (Switzerland)*, vol. 12, no. 11, pp. 1–28, 2020, doi: 10.3390/w12113223.
24. J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, “A comparison of random forest variable selection methods for classification prediction modeling,” *Expert Syst. Appl.*, vol. 134, pp. 93–101, 2019, doi: 10.1016/j.eswa.2019.05.028.
25. C. C. Aggarwal, *Neural Networks and Deep Learning*. Cham: Springer International Publishing, 2023. doi: 10.1007/978-3-031-29642-0.
26. A. K. Yadav and S. S. Chandel, “Solar radiation prediction using Artificial Neural Network techniques: A review,” *Renew. Sustain. Energy Rev.*, vol. 33, pp. 772–781, 2014, doi: 10.1016/j.rser.2013.08.055.
27. Y. chen Wu and J. wen Feng, “Development and Application of Artificial Neural Network,” *Wirel. Pers. Commun.*, vol. 102, no. 2, pp. 1645–1656, 2018, doi: 10.1007/s11277-017-5224-x.
28. Y. Jia, S. Jin, P. Savi, Q. Yan, and W. Li, “Modeling and theoretical analysis of gnss-r soil moisture retrieval based on the random forest and support vector machine learning approach,” *Remote Sens.*, vol. 12, no. 22, pp. 1–24, 2020, doi: 10.3390/rs12223679.
29. W. Zhao, A. Li, P. Huang, H. Juclin, and M. Xianming, “Surface soil moisture relationship model construction based on random forest method,” *Int. Geosci. Remote Sens. Symp.*, vol. 2017-July, pp. 2019–2022, 2017, doi: 10.1109/IGARSS.2017.8127378.
30. dan G. (BMKG) Badan Meteorologi, Klimatologi, “Waspada Kemarau! BMKG Sebut Indonesia Berpotensi Alami Kekeringan Meteorologis.” Accessed: Jun. 07, 2024. [Online]. Available: <https://www.bmkg.go.id/press-release/?p=waspada-kemarau-bmkg-sebut-indonesia-berpotensi-alami-kekeringan-meteorologis&tag=press-release&lang=ID>
31. Schölkopf, B., K. Sung, J.C. Chris, C. Burges, F. Girosi, T. Poggio, and V. Vapnik, 1997. Comparing Support Vector Machines With Gaussian Kernels to Radial Basis Function Classifiers.” *IEEE Transactions on Signal Processing*, 45(11):2758-2765.