



Downscaling spasial data curah hujan TRMM di wilayah Sumatera Selatan dengan algoritma *artificial neural network* dan *random forest regression*

HARISDIANTO¹, AZHAR K. AFFANDI^{1*}, MENIK ARIANI¹, DAN SUHADI²

¹ Jurusan Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sriwijaya. ² Jurusan Pendidikan Fisika, Fakultas Tarbiyah dan Keguruan, Universitas Islam Negeri Raden Fatah Palembang

Kata kunci:

tropical rainfall measuring mission;
machine learning;
artificial neural network;
random forest regression

ABSTRAK: Pemahaman tentang karakter distribusi curah hujan temporal dan spasial sangat penting dalam menunjang studi tentang ekologi, meteorologi, dan hidrologi. Penelitian ini memetakan kembali distribusi spasial curah hujan tahunan TRMM di wilayah Sumatera Selatan pada resolusi tinggi menggunakan teknik *downscaling* berbasis *machine learning*, *Artificial Neural Network* (ANN) dan *Random Forest Regression* (RFR). Prediktor yang digunakan yaitu NDVI, DEM, Longitude dan Latitude. Proses *downscaling* spasial curah hujan TRMM dengan model ANN memiliki akurasi R^2 0,6494, RMSE 728 mm/tahun dan MAE 715 mm/tahun. Model RFR memiliki kinerja lebih baik dengan nilai R^2 0,6818, RMSE 695 mm/tahun dan MAE 683 mm/tahun.

Keywords:

tropical rainfall measuring mission;
machine learning;
artificial neural network;
random forest regression

ABSTRACT: Understanding of the spatial and temporal precipitation characteristics is the significant importance to research fields such as ecology, meteorology, and hydrology. This study reprojected to high resolution the spatial distribution of TRMM annual precipitation over region of the South Sumatera using downscaling techniques based on machine learning, Artificial Neural Network (ANN) and Random Forest Regression (RFR). The predictors are NDVI, DEM, Longitude and Latitude. The spatial downscaling process of TRMM precipitation with the ANN model has an accuracy of R^2 0.6494, RMSE 728 mm/year and MAE 715 mm/year. The RFR model has better performance with an accuracy of R^2 0.6818, RMSE 695 mm/year and MAE 683 mm/year.

1 PENDAHULUAN

Curah hujan merupakan faktor penting dalam siklus air global dan menjadi parameter utama dari studi tentang ekologi, meteorologi dan hidrologi [1–6]. Curah hujan juga memberikan dampak yang signifikan pada kondisi agrikultur, ekonomi dan lingkungan [7,8]. Oleh karena itu, perlu pemahaman tentang karakter distribusi curah hujan secara temporal dan spasial [5]. Pengamatan langsung menggunakan stasiun penakar hujan memberikan akurasi tinggi disatu titik pengukuran, namun sulit mengekstrapolasi secara spasial untuk mendapatkan akurasi pemetaan curah hujan pada resolusi tinggi [9,10]. Sebagai alternatifnya pengamatan curah hujan dapat menggunakan data citra satelit [8,9].

Pemetaan spasial curah hujan menggunakan data satelit telah digunakan secara luas, seperti data satelit *Tropical Rainfall Measuring Mission* (TRMM)

[9–11]. TRMM adalah satelit observasi curah hujan di wilayah tropis yang meliputi sebagian besar wilayah Indonesia. TRMM memiliki skala grid $0,25^\circ \times 0,25^\circ$ pada rekam pengukuran 3 jam-an [12]. Meskipun memiliki data yang konsisten, namun skala gridnya masih tinggi sehingga sulit untuk mendapatkan informasi secara detail pada cakupan wilayah relatif lebih kecil. Maka dari itu diperlukan teknik *downscaling* untuk menurunkan skala grid supaya resolusinya meningkat.

Konsep dasar teknik *downscaling* adalah mencari model hubungan korelasi antara curah hujan terhadap variabel lingkungan sebagai prediktor pada skala grid besar, kemudian menggunakan variabel lingkungan pada skala grid kecil sebagai variabel *input*-an untuk merunkan skala grid curah hujan dari resolusi rendah menjadi resolusi tinggi [9]. Pemilihan variabel prediktor menjadi sangat penting dalam penentuan model dan mempengaruhi hasil akhir proses *downscaling*. Korelasi positif antara curah

* Corresponding Author: email: azharka@unsri.ac.id

hujan dengan variabel lingkungan seperti *Normalized Difference Vegetation Index* (NDVI) dan *Digital Elevation Model* (DEM) menjadi prediktor yang sering digunakan dalam teknik *downscaling* [9–11]. Selain variabel indek vegetasi dan elevasi, Abdelrazek Elnashar dkk. melibatkan variabel geospasial (Longitude dan Latitude) sebagai variabel prediktor [9].

Selain itu, pemilihan algoritma *downscaling* juga sama pentingnya dalam menentukan akurasi hasil akhir. Kondisi wilayah yang heterogen, dan hubungan nonlinier yang kompleks variasi spasial terhadap curah hujan membuat tidak mudah untuk memetakan curah hujan berbasis data satelit menggunakan algoritma regresi statistik tradisional [14]. Sebagai solusinya, algoritma *machine learning* dapat menyelesaikan masalah yang kompleks tersebut.

Tujuan penelitian ini menerapkan teknik *downscaling* di wilayah Sumatera Selatan menggunakan algoritma *Artificial Neural Network* (ANN) dan *Random Forest Regression* (RFR), untuk mendapatkan peta distribusi curah hujan tahunan (2019) pada skala grid 1 km.

2 METODOLOGI

Lokasi Penelitian

Lokasi data penelitian mencakup area Provinsi Sumatera Selatan (Gambar 1). Secara geografis Sumatera Selatan terletak antara 1° - 4°LS dan 102° - 106°BT, dengan luas wilayah 91.806,36 km² [15]. Topografi yang bervariasi dari Timur ke Barat dengan ketinggian rata-rata antara 400 – 1700 meter di atas permukaan laut (mdpl).

Data

Sumber data yang digunakan yaitu data spasial citra satelit dari *Earth Engine Data*, yang difilter berdasarkan tahun 2019 dan dipotong berdasarkan batas administrasi Provinsi Sumatera Selatan (Tabel 1). Selain data citra satelit, penelitian ini juga menggunakan data 11 Stasiun *Rain Gauge* BMKG untuk validasi hasil *downscaling*. Persiapan data menggunakan *Google Earth Engine* (GEE) berbasis JAVA dan pengolahan data dilakukan menggunakan *Google Colaboratory* (Colab) berbasis Python.

Tabel 1. Data Citra Satelit

No.	Data	Resolusi	Satelit
1.	Curah Hujan TRMM	27830 Meter	TRMM/3B43V7
2.	<i>Normalized Difference Vegetation Index</i> (NDVI)	1000 Meter	MODIS/MOD13A2
3.	<i>Digital Elevation Model</i> (DEM)	90 Meter	CGIAR/SRTM90_V4
4.	<i>Land Cover</i> (LC)	500 Meter	MODIS/MCD12Q1

Metode

Downscaling

Downscaling telah diperkenalkan oleh Von Storch, Zorita, dan Cubasch (1993) yang merujuk pada pendekatan statistik untuk mengubah grid berskala besar menjadi grid berskala kecil. Proses *downscaling* bisa diekspresikan seperti persamaan (1) berikut;

$$P = F(X_i) + \varepsilon \quad (1)$$

Dimana P adalah curah hujan *downscaling*, X_i adalah variabel prediktor, ε adalah residu (*error*). Ilustrasi *downscaling* ditunjukkan pada Gambar 2.

Artificial Neural Network (ANN)

ANN adalah algoritma pemrograman yang diterapkan pada proses komputasi komputer dengan mengadaptasi struktur dan karakteristik jaringan syaraf biologis manusia [18]. Arsitektur ANN terdiri dari beberapa lapisan (*hidden layers*) dan neuron seperti ditunjukkan pada Gambar 3.

Random Forest Regression (RFR)

Menurut Breiman (2001), *Random Forest* adalah sebuah algoritma yang menggabungkan beberapa pohon keputusan yang dibuat secara acak untuk menghasilkan prediksi yang lebih akurat. Arsitektur RF seperti pada Gambar 4.

Uji Parameter

Indikator parameter baik model *machine learning* dan hasil *downscaling* ditentukan berdasarkan performa model dan korelasi antara data curah hujan BMKG dengan data curah hujan hasil *downscaling*, menggunakan koefisien determinasi (R^2), *Root Mean Square Error* (RMSE), dan *Mean Absolute Error* (MAE) seperti pada persamaan (2), (3), dan (4).

$$R^2 = \frac{\sum_{i=1}^n [(S_i - S)(P_i - P)]}{\sqrt{\sum_{i=1}^n (S_i - S)^2} \sqrt{\sum_{i=1}^n (P_i - P)^2}} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (S_i - P_i)^2}{n}} \quad (3)$$

$$MAE = \frac{\sum_{i=1}^n |S_i - P_i|}{n} \quad (4)$$

dimana S adalah curah hujan TRMM dan P adalah prediksi curah hujan TRMM. R^2 menunjukkan seberapa kuat hubungan antara S dan P . RMSE adalah kesalahan suatu model dalam memprediksi data dan MAE sebagai indikator bias.

Prosedur Penelitian

Secara umum penelitian ini terbagi dalam empat tahapan, yaitu persiapan data, optimasi model prediksi, menghitung residu, dan proses *downscaling*.

Persiapan Data

Data dipersiapkan dalam 2 skala grid yaitu 25 km dan 1 km. Data curah hujan TRMM, NDVI dan DEM diproyeksi ulang pada skala grid 25 km menggunakan teknik tetangga terdekat (*nearest neighbor technique*) dan diagregat pada tahun 2019 (Gambar 5) [9]. Data pada skala grid 1 km dipersiapkan dari data NDVI dan proyeksi ulang dari data DEM dan LC (Gambar 6). Semua data di-export dan disimpan dalam format *GeoTIFF* atau *raster*.

Optimasi Model Prediksi

Data dinormalisasi menggunakan *MinMaxScaler* kemudian dibagi dalam 2 bagian, yaitu 90% untuk data latih (*training*) dan 10% untuk data uji (*testing*). Penentuan *Hyper-parameter* model seperti terlihat pada Tabel 2. Model digunakan untuk prediksi curah hujan TRMM pada skala grid 25 km, kemudian dievaluasi dengan nilai tertinggi dari R^2 , dan nilai terkecil dari RMSE dan MAE.

Tabel 2. Hyper-parameter model

Machine Learning	Hyper-parameter
RFR	max_depth = 7, n_estimators = 150
ANN	layer = 6, units = 100, input_dim = 4, activation = 'relu', batch_size = 32, epochs = 100, optimizer = 'adam'

Menghitung Residu

Tahap ini adalah menghitung residu pada pada sakal grid 25 km menggunakan persamaan (5), kemudian diinterpolasi secara linier ke skala grid 1 km.

$$\varepsilon_{25\text{ km}} = TRMM_{25\text{ km}} - TRMM_{predicted\ 25\text{ km}} \quad (5)$$

Proses Downscaling

Nilai negatif dari variabel prediktor NDVI indikasi area perkotaan dan area pemukiman yang minim vegetasi dan akan mempengaruhi proses *downscaling*, sehingga nilai negatif dieliminasi berdasarkan data LC [9]. Proses *downscaling* dilakukan menggunakan model optimasi pada poin 2 diatas, namun variabel prediktor pada skala grid 1 km. Hasil *downscaling* dikoreksi dengan menambahkan residu pada skala grid 1 km dari point 3, kemudian divalidasi menggunakan data stasiun BMKG.

3 HASIL DAN PEMBAHASAN

Optimasi Model Prediksi

Model yang dibangun memiliki korelasi sangat baik antara data TRMM dan curah hujan hasil prediksi (Gambar 7). Akurasi kedua model memiliki akurasi tinggi yaitu diatas 92%. Hal tersebut dapat dilihat dari nilai R^2 0,9287 pada ANN dan 0,9332 pada RFR. Sedangkan kesalahan (*error*) yang cukup rendah yaitu RMSE 86 mm/tahun dan MAE 56 mm/tahun pada ANN, kemudian RSME 83 mm/tahun dan MAE 59 mm/tahun pada RFR.

Interpretasi model penting untuk melihat kontribusi setiap variabel prediktor dalam membangun sebuah model *machine learning* [19], [20]. Penelitian ini menggunakan *SHapley Additive exPlanation* (SHAP) untuk investigasi pengaruh dominan variabel prediktor terhadap model. Gambar 8 adalah grafik SHAP yang menunjukkan kontribusi setiap variabel terhadap model. Longitude dan Latitude merupakan variabel dominan pada model ANN, hal serupa juga ditunjukkan oleh penelitian Abdelrazek Elnashar dkk., 2020 [9]. Sedangkan pada model RFR variabel DEM lebih signifikan berkontribusi terhadap model, ini juga dijumpai pada penelitian oleh Irfan Alghani Khalid dan Imas Sukaesih Sitanggang, 2022 [21].

Downscaling 1 km dan Validasi

Hasil *downscaling* menggunakan model ANN ditunjukkan pada Gambar 9(a) dan model RFR Gambar 9(b). Distribusi spasial curah hujan dari kedua model sama-sama menghasilkan curah hujan rendah dengan curah hujan ≤ 2200 mm/tahun dominan dibagian timur Sumatera Selatan dan curah hujan tinggi dengan curah hujan ≥ 2400 mm/tahun dibagian baratnya. Visualisasi model ANN lebih halus, sedangkan model RFR terlihat lebih kontras, dan juga terlihat adanya efek bayangan grid 25 km. Area putih pada peta merupakan area perkotaan dan pemukiman.

Validasi curah hujan hasil *downscaling* dengan data stasiun BMKG menunjukkan model RFR memiliki akurasi lebih baik dari ANN, terlihat dari nilai R^2 yaitu 0,6818, RMSE 695 mm/tahun dan MAE 683 mm/tahun untuk RFR, dan ANN yaitu R^2 0,6494, RMSE 728 mm/tahun dan MAE 715 mm/tahun (Gambar 10).

4 KESIMPULAN

Penelitian ini berhasil melakukan *downscaling* data curah hujan TRMM menggunakan algoritma *machine learning* (ANN dan RFR), dengan prediktor index vegetasi, elevasi, dan geospasial. Optimasi model yang dibangun memiliki performa akurasi diatas 92% dalam memprediksi data curah hujan TRMM pada skala grid 25 km. Meskipun demikian didapat hasil yang berbeda pada model RFR, hasil akhir *downscaling* terdapat efek bayangan variabel geospasial dari skala grid 25 km. Validasi antara data curah hujan hasil *downscaling* algoritma ANN dan RFR dengan data observasi BMKG (Januari – Oktober 2019) menghasilkan nilai akurasi $R^2 > 64\%$, nilai tersebut belum optimal dikarenakan data observasi BMKG yang dimiliki tidak genap 1 tahun observasi. Model *machine learning* dan kombinasi prediktor lain dapat diterapkan pada proses *downscaling* untuk mendapatkan nilai akurasi lebih tinggi.

REFERENSI

- [1] S. Chen, L. Zhang, D. She, and J. Chen, "Spatial downscaling of Tropical Rainfall Measuring Mission (TRMM) annual and monthly precipitation data over the middle and lower reaches of the Yangtze River Basin, China," *Water (Switzerland)*, vol. 11, no. 3, 2019, doi: 10.3390/w11030568.
- [2] T. Zhang *et al.*, "Spatial downscaling of TRMM precipitation data considering the impacts of macro-geographical factors and local elevation in the Three-River Headwaters Region," *Remote Sens. Environ.*, vol. 215, no. June, pp. 109–127, 2018, doi: 10.1016/j.rse.2018.06.004.
- [3] W. Jing, Y. Yang, X. Yue, and X. Zhao, "A spatial downscaling algorithm for satellite-based precipitation over the Tibetan plateau based on NDVI, DEM, and land surface temperature," *Remote Sens.*, vol. 8, no. 8, 2016, doi: 10.3390/rs8080655.
- [4] Y. Shi *et al.*, "Mapping annual precipitation across Mainland China in the period 2001-2010 from TRMM3B43 product using spatial downscaling approach," *Remote Sens.*, vol. 7, no. 5, pp. 5849–5878, 2015, doi: 10.3390/rs70505849.
- [5] S. Jia, W. Zhu, L. Aifeng, and T. Yan, "Spatial downscaling algorithm of TRMM precipitation based on NDVI and DEM in the Qaidam Basin of China," *Remote Sens. Environ.*, vol. 115, no. 12, pp. 3069–3079, 2011, doi: 10.1016/j.rse.2011.06.009.
- [6] W. W. Immerzeel, M. M. Rutten, and P. Droogers, "Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula," *Remote Sens. Environ.*, vol. 113, no. 2, pp. 362–370, 2009, doi: 10.1016/j.rse.2008.10.004.
- [7] M. Irfan, W. Mardiansyah, M. Ariani, A. Sulaiman, and I. Iskandar, "Is TRMM product good proxy for gauge precipitation over peat land area of the South Sumatra?," *J. Phys. Conf. Ser.*, vol. 1282, no. 1, 2019, doi: 10.1088/1742-6596/1282/1/012021.
- [8] Z. Duan and W. G. M. Bastiaanssen, "First results from Version 7 TRMM 3B43 precipitation product in combination with a new downscaling-calibration procedure," *Remote Sens. Environ.*, vol. 131, pp. 1–13, 2013, doi: 10.1016/j.rse.2012.12.002.
- [9] A. Elnashar *et al.*, "Downscaling TRMM monthly precipitation using google earth engine and google cloud computing," *Remote Sens.*, vol. 12, no. 23, pp. 1–22, 2020, doi: 10.3390/rs12233860.
- [10] Y. Zhang, Y. Li, X. Ji, X. Luo, and X. Li, "Fine-resolution precipitation mapping in a mountainous watershed: Geostatistical downscaling of TRMM products based on environmental variables," *Remote Sens.*, vol. 10, no. 1, pp. 1–27, 2018, doi: 10.3390/rs10010119.
- [11] Z. Ma, Z. Shi, Y. Zhou, J. Xu, W. Yu, and Y. Yang, "A spatial data mining algorithm for downscaling TMPA 3B43 V7 data over the Qinghai-Tibet Plateau with the effects of systematic anomalies removed," *Remote Sens. Environ.*, vol. 200, no. December 2016, pp. 378–395, 2017, doi: 10.1016/j.rse.2017.08.023.
- [12] G. J. Huffman *et al.*, "The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales," *J. Hydrometeorol.*, vol. 8, no. 1, pp. 38–55, 2007, doi: 10.1175/JHM560.1.
- [13] X. Zhao, W. Jing, and P. Zhang, "Mapping fine spatial resolution precipitation from trmm precipitation datasets using an ensemble learning method and modis optical products in china," *Sustain.*, vol. 9, no. 10, pp. 1–17, 2017, doi: 10.3390/su9101912.
- [14] W. Jing, Y. Yang, X. Yue, and X. Zhao, "A comparison of different regression algorithms for downscaling monthly satellite-based precipitation over North China," *Remote Sens.*, vol. 8, no. 10, pp. 1–17, 2016, doi: 10.3390/rs8100835.
- [15] Bappeda Sumsel, "RPJMD Prov Sumsel 2013-2018," 2018, pp. 1–134.
- [16] H. Von Storch, E. Zorita, and U. Cubasch, "Downscaling of Global Climate Change Estimates to Regional Scales: An Application to Iberian Rainfall in Wintertime," *J. Clim.*, vol. 6, no. 6, pp. 1161–1171, 1993.
- [17] N. W. Park, Y. Kim, and G. H. Kwak, "An Overview of Theoretical and Practical Issues in Spatial Downscaling of Coarse Resolution Satellite-derived Products," *Ko-*

rean *J. Remote Sens.*, vol. 35, no. 4, pp. 589–607, 2019, doi: 10.7780/kjrs.2019.35.4.8.

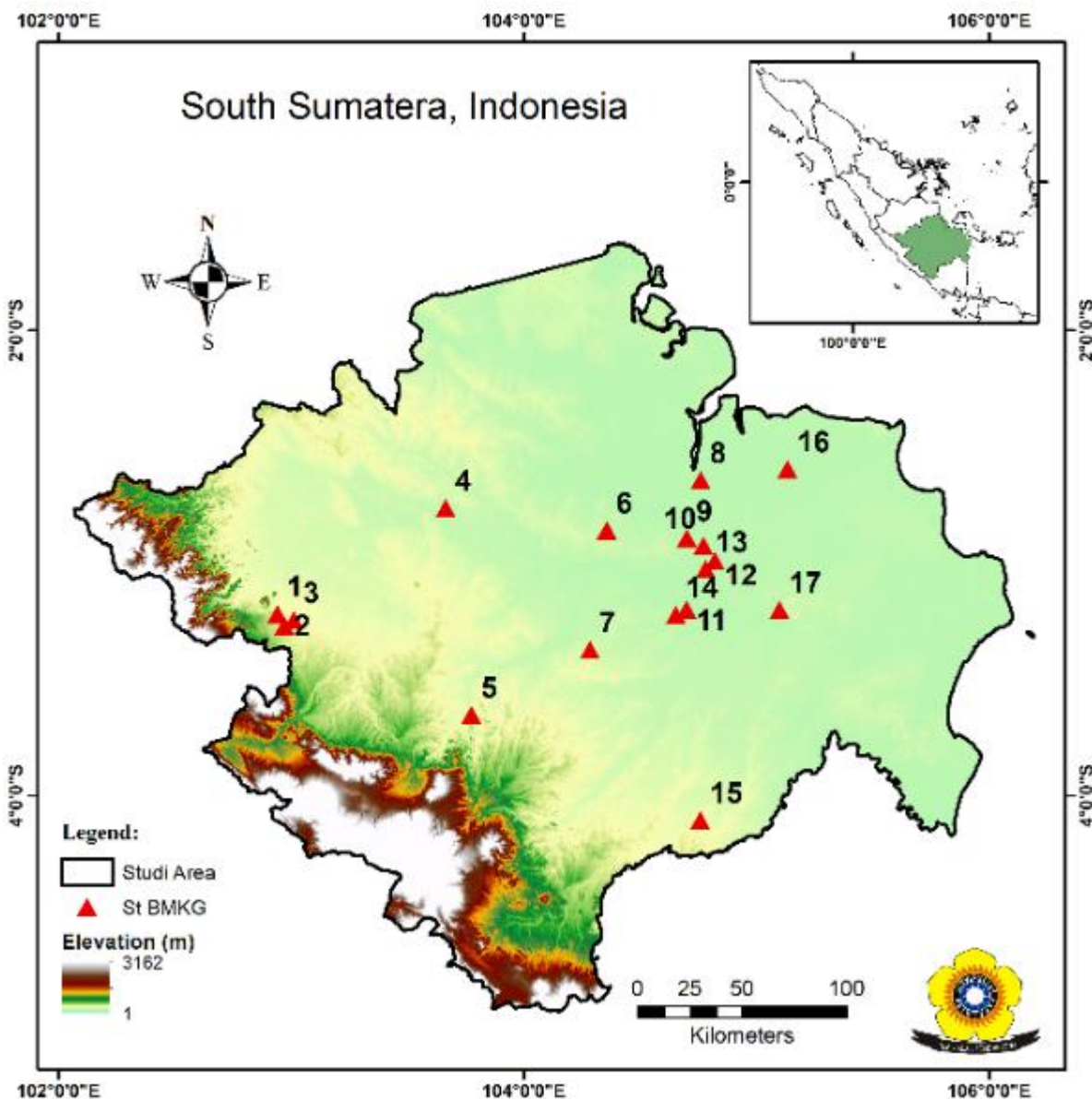
[18] L. V. Fausett, *Fundamentals of Neural Networks: Architectures, Algorithms And Applications*. 1994. doi: 10.1109/T-C.1969.222718.

[19] S. M. Lundberg, G. G. Erion, and S.-I. Lee, “Consistent Individualized Feature Attribution for Tree Ensembles,” 2018, vol. 03888, no. 2018, p. 3888. [Online]. Available: <http://arxiv.org/abs/1802.03888>

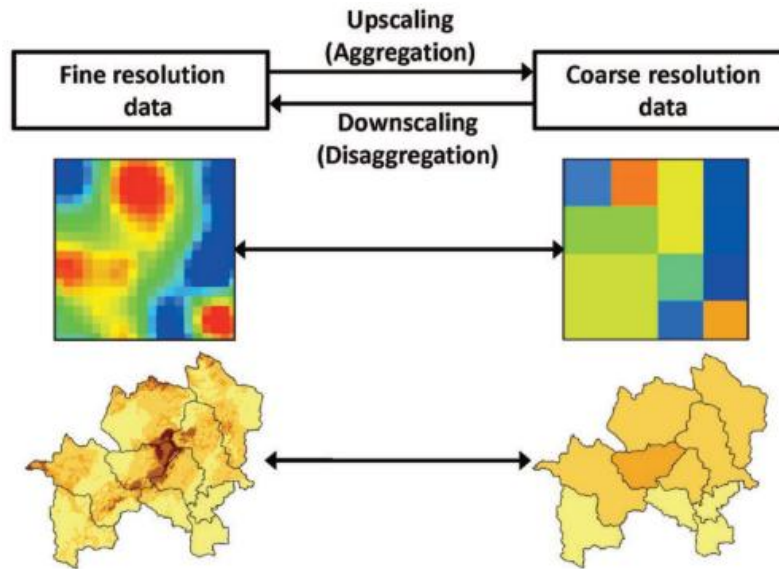
[20] K. K. P. M. Kannangara, W. Zhou, Z. Ding, and Z. Hong, “Investigation of feature contribution to shield tunneling-induced settlement using Shapley additive explanations method,” *J. Rock Mech. Geotech. Eng.*, vol. 14, no. 4, pp. 1052–1063, 2022, doi: 10.1016/j.jrmge.2022.01.002.

[21] I. A. Khalid and S. Sitanggang, “Machine Learning-Based Spatial Downscaling on Precipitation Satellite Data in Riau Province, Indonesia,” *Turkish J. Comput. Math. Educ.*, vol. 13, no. 02, pp. 7–16, 2022, [Online]. Available: <https://gadm.org/data.html>

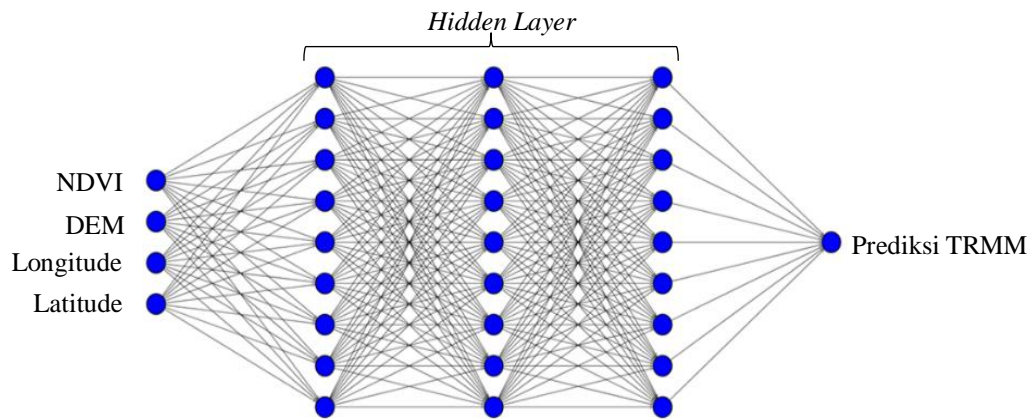
LAMPIRAN



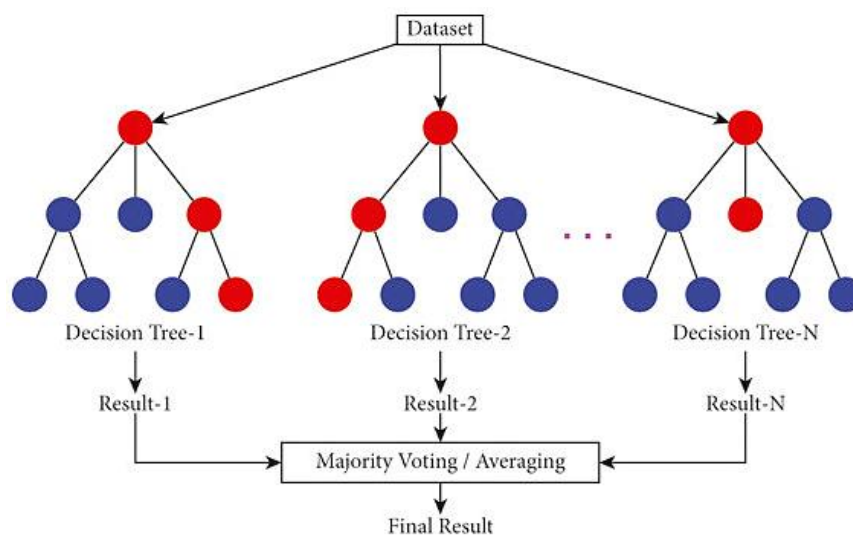
Gambar 1. Peta Lokasi penelitian, DEM, dan Stasiun BMKG



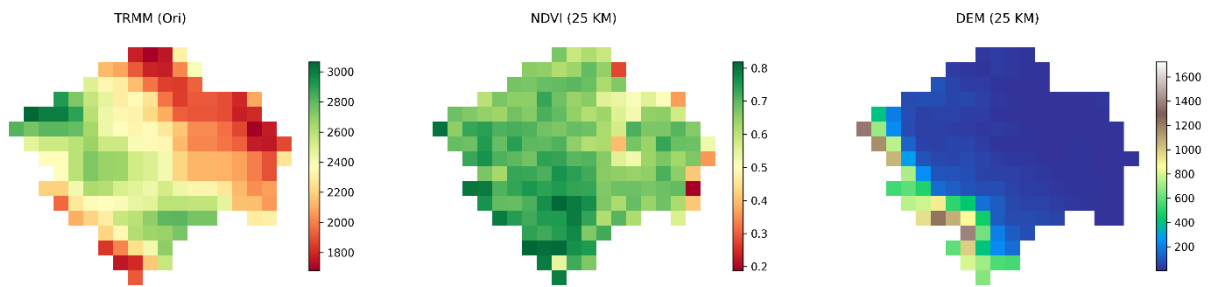
Gambar 2. Ilustrasi *downscaling* spasial [17].



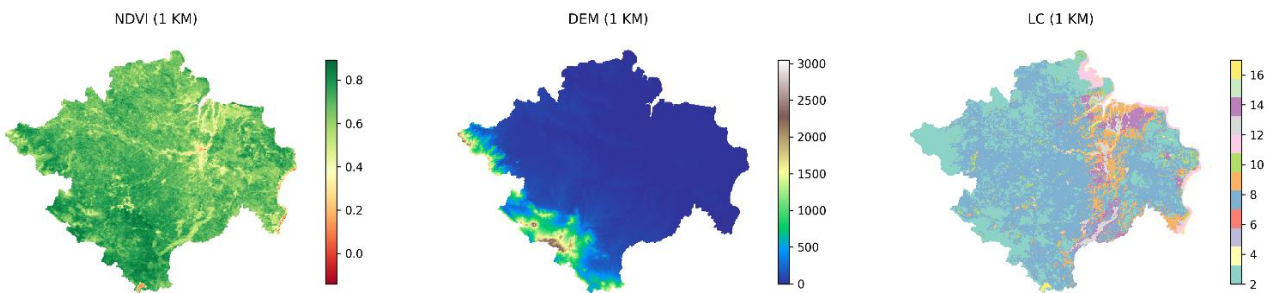
Gambar 3. Arsitektur ANN [9].



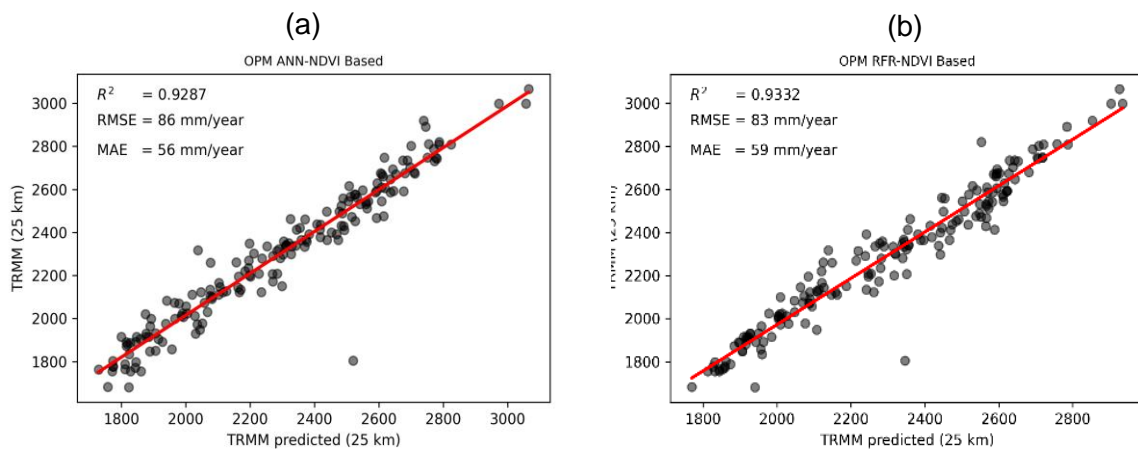
Gambar 4. Arsitektur *Random Forest* (sumber: researchgate.net).



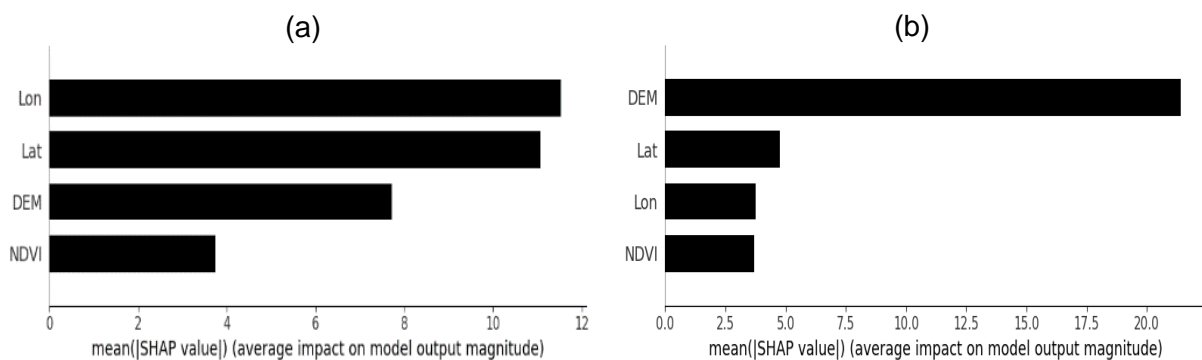
Gambar 5. Data satelit skala grid 25 km



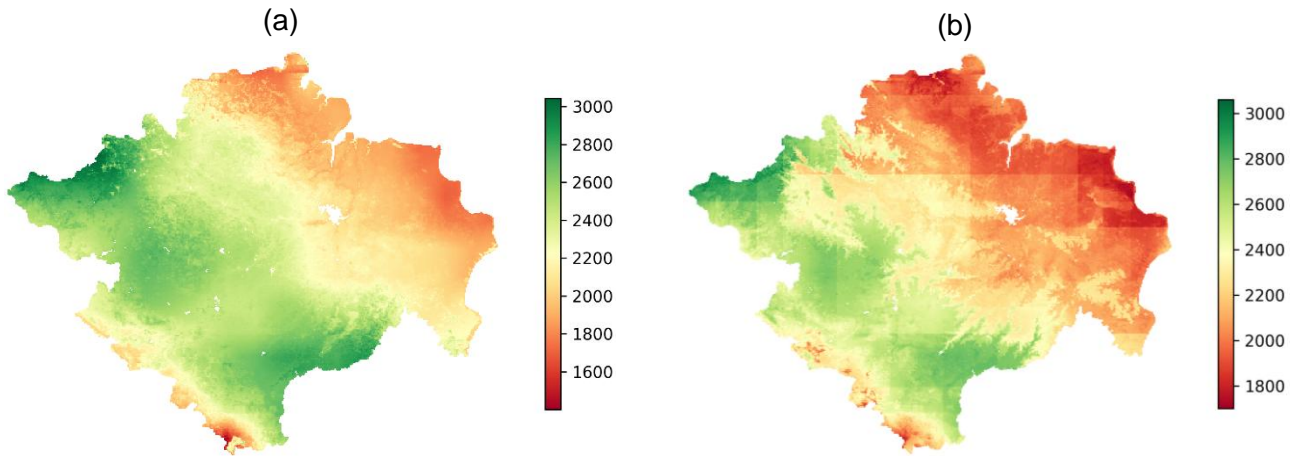
Gambar 6. Data satelit skala grid 1 km



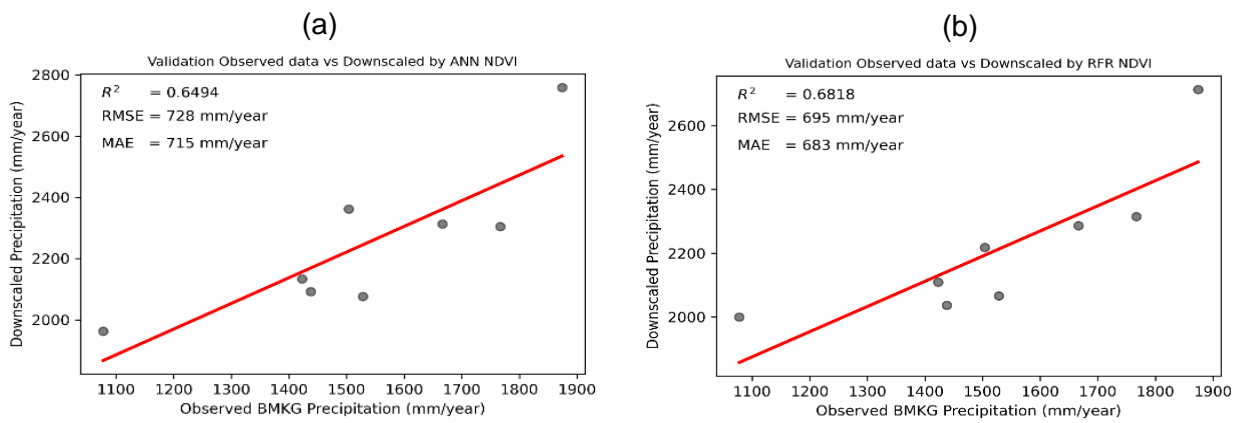
Gambar 7. Korelasi curah hujan TRMM dengan prediksi curah hujan TRMM pada skala grid 25 km, model ANN (a) dan model RFR (b)



Gambar 8. Grafik SHAP, ANN (a) dan RFR (b).



Gambar 9. Hasil *downscaling* pada skala grid 1 km, ANN (a) dan RFR (b)



Gambar 10. Validasi curah hujan hasil *downscaling* dengan curah hujan stasiun BMKG, ANN (a) dan RFR (b).