Research Article

Machine Learning-Based Spatial Downscaling on Precipitation Satellite Data in Riau Province, Indonesia

Irfan Alghani Khalid [a], Imas Sukaesih Sitanggang[a]

[a] Department of Computer Science, IPB University, Indonesia

Abstract: Precipitation is a condensation process on water from the atmosphere that causes rain. Rain has positive and negative impacts on society, especially because of global warming. Estimating precipitation at specific locations is needed to mitigate the negative consequences. Utilizing satellite data is the best way to estimate precipitation, but it has a coarse resolution. Therefore, a study was conducted on spatial downscaling for increasing spatial resolution of precipitation data from the tropical rainfall measuring mission satellite using machine learning in Riau Province, Indonesia. We compared machine learning models namelydecision trees, multiple linear regression, support vector machines, and random forest to downscale the data. We used variables like normalized difference vegetation index, digital elevation model, and land cover as the input for the model. Also, we validated the result with the measurement of the rain gauge station at Riau Province Indonesia. Based on the study, we found that the decision tree is the best model to downscale the precipitation data in Riau Province with the mean square error value of 0.00048 and the R2 value of 0.67107. We found that the digital elevation model is the most important variable in downscaling the precipitation data.

Keywords: Big data, machine learning, precipitation, remote sensing, spatial downscaling.

1. Introduction

Precipitation is a condensation process on water from the atmosphere. This process has caused the water from the atmosphere to fall to the earth's surface. This falling water is called rain. Rain has so many positive impacts on all creatures. But there are also negative impacts because of global warming. A study conducted by **Tabari (2020)** showed that the global warming process causes a higher intensity of bad weather throughout the region. The data from NOAA (https://www.epa.gov/climate-indicators/climate-change-indicators-heavy-precipitation) also showed that the percentage of precipitation is increasing each year by around 0.5 percent. This impact is causing phenomena like crop failure, ground erosion, and flood. Those phenomena can affect a tremendous loss for society **(Bell et al., 2016)**. With the increasing intensity of precipitation, a method for calculating the precipitation intensity is needed.

There are two methods for retrieving the precipitation data. The first method is to retrieve the data from the rain gauge station. The advantage of this method is it can calculate the precipitation value accurately. But the disadvantage of this method is the amount of data is limited. In other words, some locations cannot be calculated **(Zhan et al., 2018)**. The second method is to gather the data from the satellite. Tropical Rainfall Measuring Mission (TRMM) is one of the satellites that measure precipitation. The satellite can capture precipitation with a spatial resolution of 25 KM. It means that one pixel of the image captures information with an area size of 25 kilometers. The advantage of this method is that the satellite can measure precipitation values throughout the all-region. Because of the coarse resolution, the satellite cannot measure the precipitation precisely **(Mahmud et al., 2018)**. For retrieving the precise value, a method for increasing the spatial resolution is needed. The method is called spatial downscaling. This studyproposes to apply the machine-learning-based spatial downscaling method for measuring precipitation in Riau Province, Indonesia. The final output from the research is to gather and validate the final map with the spatial resolution by 1 km.

2.Review of Related Studies

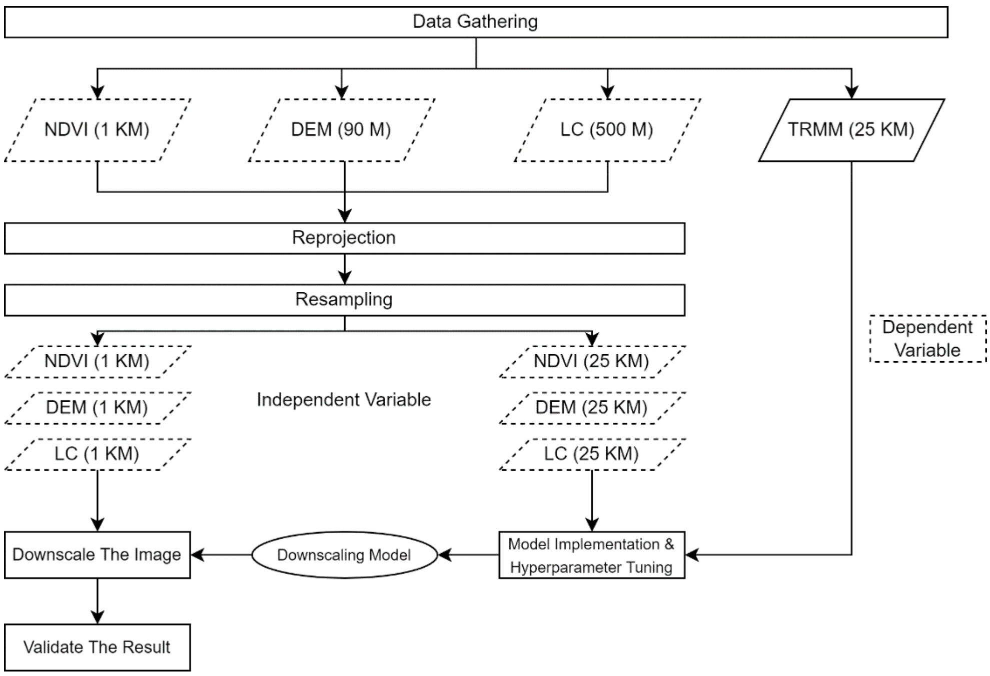
Jing et al., (2016) implemented a downscaling algorithm to downscale the precipitation data from the TRMM satellite. The machine learning model like Random Forests (RF), Support Vector Machine (SVM), and Exponential Regression (ER) were compared. Those models used variables Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST), and the Digital Elevation Model (DEM). The experiment showed that the SVM model outperformed all machine learning models when the inputs were the NDVI and DEM. The RF model outperformed all models when the LST was included in the model.**Liu et al., (2018)** used a machine learning model to downscale the soil moisture data. The experiment used data includingLand Surface Temperature (LST) and Digital Elevation Model (DEM) to learn associations between them and the soil moisture data. From the experiment, the k-Nearest Neighbours (KNN) model was the best model to downscale the data.**Elnashar et al., (2020)** compared Gradient Boosting (GB), Support Vector Machine (SVM), and Artificial Neural Network (ANN) to downscale the precipitation data. Vegetation index, topography, and geolocation were used to train the model. The experiment showed the ANN model outperformed all algorithms to downscale the precipitation data.**Yan & Bai, (2020)** compared RF, SVM, and KNN models to downscale the soil moisture data. Those models took vegetation index, land cover, topography, and LST data to learn patterns with the soil moisture data. The experiment showed the RF model outperformed all machine learning models.**Xu et al., (2021)** proposed a downscaling model that combines the RF model and kriging to downscale the LST data. The model learned patterns between the Sentinel-2A data and the LST data. The proposed model was compared with an NDVI-based model called Thermal Sharpening Method (TsHARP). The result showed that the proposed model outperforms the TsHARP model. **Lezama Valdes et al., (2021)** used machine learning algorithms like Random Forest, gradient boosting, and artificial neural networks to downscale the LST data. The model took surface information and learned patterns from them with the LST variable. From the experiment, the random forest model outperformed all models that were tested.

3.Methods

3.1. Research Workflow

In general, the research steps started from gathering the data, preprocessing the data, modeling the data, to validating the result (Figure 1).

Figure.1. Research workflow.



3.2. Data Source

The data weregathered from the Google Earth Engine (GEE) platform. GEE is a cloud-based platform for analyzing spatial data, whether vector or raster data. GEE contains a computational platform for processing spatial data. It also has a data catalog that gathers all publicly available spatial data. Those datasets are ready to use, allowing to focus on specific application research **(Gorelick et al., 2017)**. To specify the research scope, we collectthe data from 2018. We picked five spatial data. Those data are the precipitation data, the Digital Elevation Model (DEM), the Normalized Difference Vegetation Index (NDVI), the land cover data, and vector data of the Riau Province border.

We retrieved precipitation datafrom the Tropical Rainfall Measuring Mission (TRMM) satellite. This satellite gathered information about precipitation in a 25 km spatial resolution from January 1998 until December 2019**(Adler et al., 2003)**.Currently, the data collection is handled by a satellite called Global Precipitation Measurement (GPM). We used the 3B43 version of the satellite to gather the annual precipitation data. Because the data was updated monthly, we aggregated it using the average method to retrieve the yearly data.

We retrieved DEM datafrom Shuttle Radar Topography Mission (SRTM). The data contains 90-meter-resolution elevation data, and it captures only once in February 2000**(Jarvis et al., 2008)**. Because there is only one elevation data, we only picked and left it without aggregating the data.

We collected NDVI datafrom MODIS with the code of MOD13A2**(Kamel, 2015)**. NDVI is an index that estimates the density of vegetated areas**(Gessesse & Melesse, 2019)**. With this sensor, we got the NDVI data without calculating it first. The data has a 500 meters resolution, and the satellite captures the data every 16 days. For getting the annual NDVI, we aggregated it using the average method.

We obtained land cover datafrom MODIS with the code MCD12Q1**(Damien, 2019)**. The data is derived from the classification result of MODIS Terra and Aqua reflectance data. It has a spatial resolution of 500 meters and is produced every year from January 2001 until January 2020. There are five different types of land cover that are captured by MODIS. This study used type 1 land cover, where the classes are based on the International Geosphere-Biosphere Programme (IGBP) proposed land cover classes. Because the data arealready yearly, we only picked the data and left it without any aggregation steps.

Table.1. Description of for each class**(Damien, 2019)**.

Value	Description
1	Evergreen Needleleaf Forests: dominated by evergreen conifer trees (canopy >2m). Tree cover >60%.
2	Evergreen Broadleaf Forests: dominated by evergreen broadleaf and palmate trees (canopy >2m). Tree cover >60%.
3	Deciduous Needleleaf Forests: dominated by deciduous needleleaf (larch) trees (canopy >2m). Tree cover >60%.
4	Deciduous Broadleaf Forests: dominated by deciduous broadleaf trees (canopy >2m). Tree cover >60%.
5	Mixed Forests: dominated by neither deciduous nor evergreen (40-60% of each) tree type (canopy >2m). Tree cover >60%.
6	Closed Shrublands: dominated by woody perennials (1-2m height) >60% cover.
7	Open Shrublands: dominated by woody perennials (1-2m height) 10-60% cover.
8	Woody Savannas: tree cover 30-60% (canopy >2m).
9	Savannas: tree cover 10-30% (canopy >2m).
10	Grasslands: dominated by herbaceous annuals (<2m).
11	Permanent Wetlands: permanently inundated lands with 30-60% water cover and >10% vegetated cover.
12	Croplands: at least 60% of area is cultivated cropland.
13	Urban and Built-up Lands: at least 30% impervious surface area including building materials, asphalt, and vehicles.
14	Cropland/Natural Vegetation Mosaics: mosaics of small-scale cultivation 40-60% with natural tree, shrub, or herbaceous vegetation.
15	Permanent Snow and Ice: at least 60% of area is covered by snow and ice for at least 10 months of the year.
16	Barren: at least 60% of area is non-vegetated barren (sand, rock, soil) areas with less than 10% vegetation.
17	Water Bodies: at least 60% of area is covered by permanent water bodies.

We retrieved land borderfrom the GADM website (https://gadm.org/data.html). GADM is a database of administrative areas from all countries. In Indonesia, there are several levels of administrative areas. For this case, we collecteddata on a province level, which in this case we took the Riau Province data.

Table.2. Satellite data description.

Product Name	Spatial Resolution	Temporal Resolution	Measurement Unit
TRMM 3B43: Monthly Precipitation Estimates	27830 meters	Monthly	mm/year
SRTM Digital Elevation Data Version 4	90 meters	-	m
MODIS Combined 16-Day NDVI	463.313 meters	16 Days	-
MODIS Land Cover Type Yearly Global 500m	500 meters	Yearly	-

3.3. Data Preparation

After we gathered the data, we prepared the data before building the downscaling model. We reprojected each of them with the reference projection. In this case, we used the EPSG:4326 projection for each data. After we reprojected the data, we resampled the data into two spatial resolutions. Those resolutions are 25 KM and 1 KM. The reason for resampling the satellite images is because each data contained different spatial resolutions. Therefore, we resampled the data into 25 KM and 1 KM for modeling and downscaling purpose, respectively. All steps were done using the GEE platform.

Figure.2. Satellite images in 25 KM resolution.

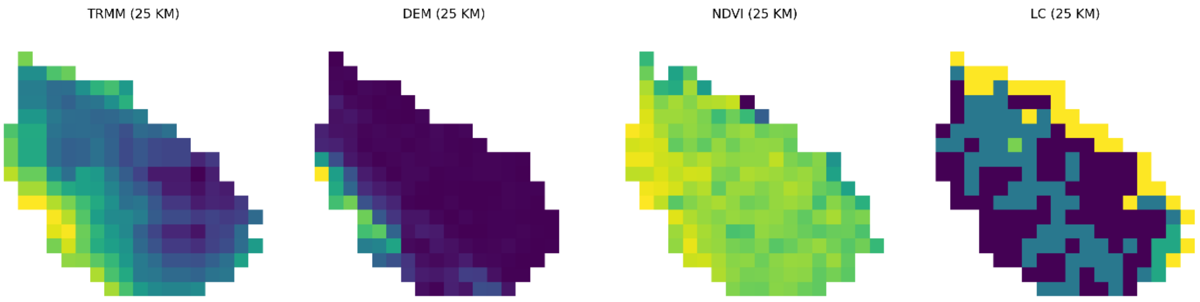
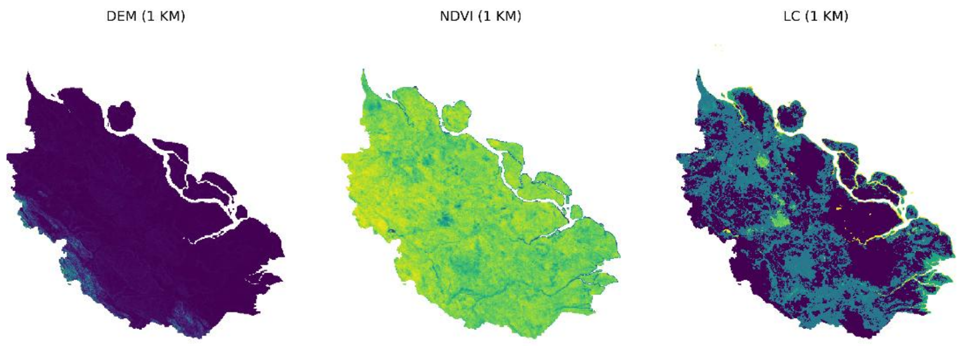


Figure.3. Resampled satellite images in 1 KM resolution.



3.4. Model Implementation

For the experiment, we compared machine learning models like Random Forest (RF), Support Vector Machine (SVM), Multiple Linear Regression (MLR), and Decision Tree (DT). We implemented all models using the scikit-learn library**(Pedregosa et al., 2011)**.Random Forest is an ensemble-based machine learning model. The algorithm generates a collection of trees, where each tree contains a random vector that is trained using a sample of the dataset. Each tree outputs a prediction result. And the final prediction will be determined using a voting method and pick the majority prediction**(Breiman, 2001)**.

Support Vector Machine (SVM) is a machine learning model that tries to create a decision vector that can separate data with different classes. The algorithm chooses a vector that has a maximum margin. The margin is calculated between the vector and the outmost data for each class. Therefore, the model can have optimal performance**(Bishop, 2006)**.

Multiple linear regression is a machine learning model that tries to learn relationships between the dependent variable y with the independent variable x. Multiple linear regression is an extension to simple linear regression where it has more than one independent variable. The model will fit a line inside the data. Therefore, the model can generalize to the unseen data.

Decision Tree is a machine learning model that segments the data into several regions. The tree is generated by adding a rule as the first node from all possible variables. Then, the model adds the branch that represents the rule recursively until the tree cannot be divided anymore. One of the advantages of this model is the interpretability of the prediction process. Rather than getting the predictions only, the model also shows the rule behind the results**(James et al., 2013)**.

3.5. Hyperparameter Tuning

The proposed model still used the default hyperparameters. Therefore, we did the hyperparameter tuning process. Hyperparameter is a parameter that is adjustable before running the model. By tuning the hyperparameters, the model can reach higher accuracy. We used the grid search method for finding the optimal hyperparameters. We used the GridSearchCV function from the scikit-learn library to tune the hyperparameters**(Pedregosa et al., 2011)**. From the hyperparameter tuning process, we chose the best machine learning model. With the best model, we used it to downscale the satellite data with the spatial resolution of 1 KM.

Table. 3. Hyperparameter combinations.

Model	Hyperparameter	Range of value
Decision Tree	Max Depth	[2, 3, 5, 10, 20]
	Minimum Samples Leaf	[5, 10, 20, 50, 100]
Support Vector Machine	C	[0.1, 1, 10, 100, 1000]
	Gamma	[1, 0.1, 0.01, 0.001, 0.0001]
	Kernel	rbf
Random Forest	Number of Estimators	[5, 20, 50, 100]
	Max Features	[auto, sqrt]
	Max Depth	[10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120]
	Minimum Samples Split	[2, 6, 10]
	Minimum Samples Leaf	[1, 3, 4]
	Bootstrap	[true, false]

To determine the best model, we used metrics like Mean Squared Error (MSE) and Coefficient of Determination (R2 score) for validating the model's performance. MSE is a metric that calculates the average of squares of errors, where the error is the difference between the predicted and the actual value. If the MSE value is closer to 0, the model has a great performance. R2 score calculates the proportion of variance from the dependent variable (y), and it tells the goodness of fit. In other words, the value measures how well the model predicts the unseen data. If the value is closer to 1, the model performs well**(James et al., 2013)**.

3.6. Result Validation

We validated the downscaled image with the data from the Indonesian Bureau of Meteorology and Geophysics (BMKG) website (http://dataonline.bmkg.go.id/home). In Riau Province Indonesia, there are three rain gauge stations, and each station locates in Pekanbaru, Kampar, and Japura. Because the data from BMKG is point-based, we sampled the downscaled result based on the coordinates from each station. We used QGIS to retrieve the downscaled result.

Table. 4. Details about the weather station.

Station's Name	Location	Latitude	Longitude	Elevation
Stasiun Meteorologi Sultan Syarif Kasim II	Pekanbaru	0.45924	101.44743	39
Stasiun Klimatologi Kampar	Kampar	0.40700	101.21700	15
Stasiun Meteorologi Japura	Japura	-0.33000	102.32000	19

4. Results

4.1. Data Exploration

Before implementing the downscaling model, we explored the dataset we used for this research. First, we observed scatter plots between input variables like NDVI and DEM with the output variable, which is the precipitation variable. Next, we analyzed the impacts of each land cover class on the precipitation value. Lastly, we checked the distributions of the precipitation value.

Figure. 4. NDVI and precipitation.

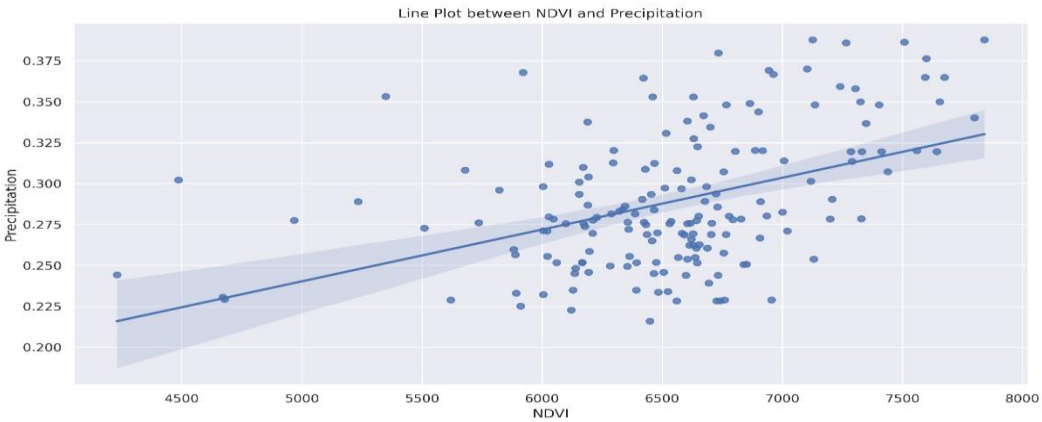
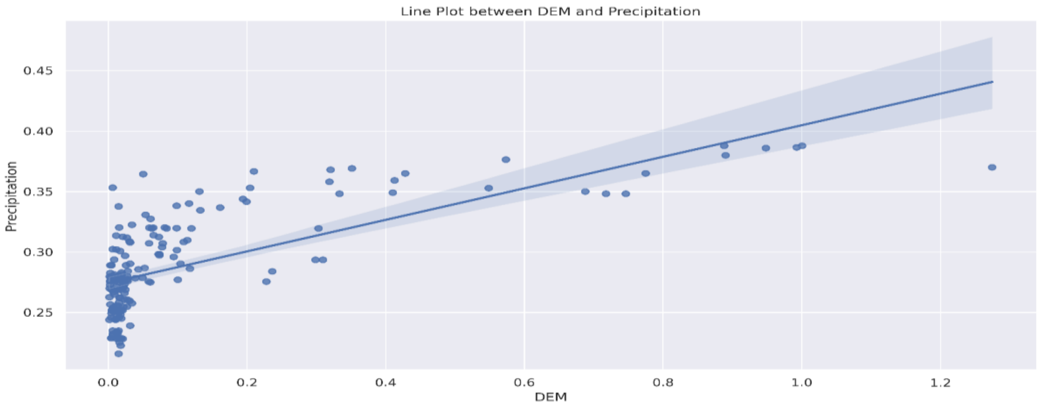


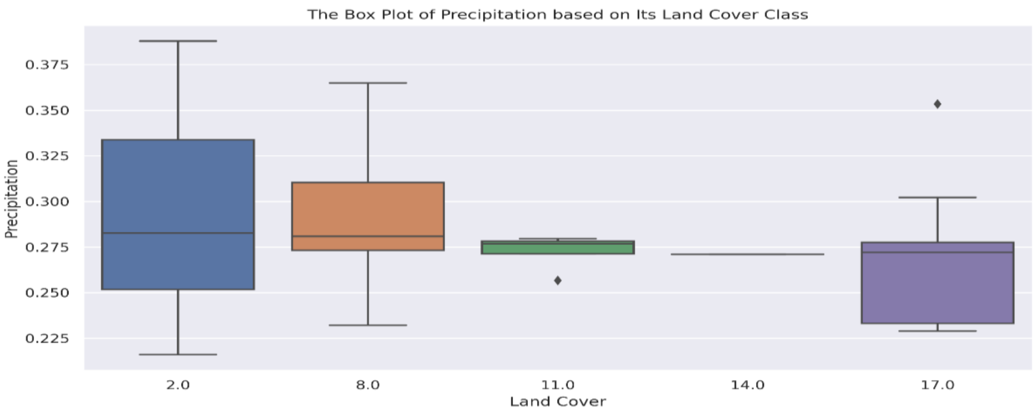
Figure. 5. DEM and precipitation.



We geta positive correlation between the NDVI variable and the precipitation variable. The scatter values are spreading but tend to have a linear relationship on both. As the number of NDVI gets higher, the precipitation value follows it.

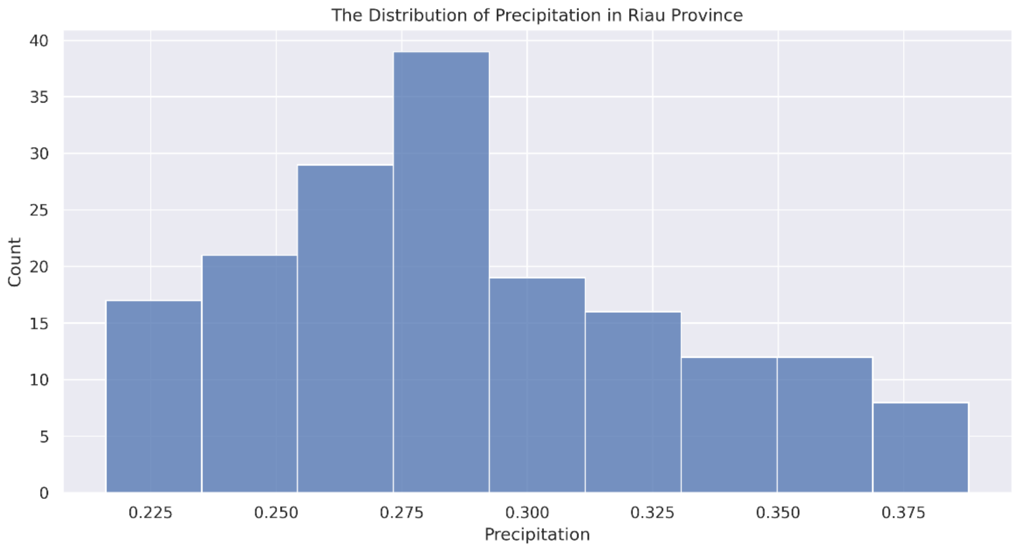
We also observed that the DEM variable is positively correlated to the precipitation variable. Although the precipitation values have a large variance when the DEM is closer to zero, the relationship between DEM and precipitation is linear as both numbers get higher. The higher the DEM value has, the higher the precipitation value becomes.

Figure. 6. Precipitation based on the land cover class.



Based on the boxplot above, we observedthat each land cover class has a different variance of precipitation value but with an almost similar median value to it. In sequence, class 2, 8, 11, 14, and 17 represents forest, savannas, wetlands, cropland, and water bodies. We foundthat class 2 tends to have a larger variance than all land cover classes. Class 8 tends to have a right-skewed distribution. Class 11 has a slightly left-skewed distribution to it while having an outlier below. Class 14 does not have any distribution and has a median value slightly below 0.275. Class 17 tends to have a left-skewed distribution while having an outlier at the above part (Figure 6).

Figure. 7. Distribution of precipitation.



Based on the histogram above, we saw that the curve shape of the precipitation's distribution tends to be normal. Because the distribution was not skewed, we did not apply any preprocessing steps to make the normal distribution.

4.2. Hyperparameter Tuning

We did the hyperparameter tuning to our proposed machine learning model, except for multiple linear regression. We used 10-fold cross validation to measure the model's performance while choosing the best hyperparameters simultaneously. Based on the process, the best hyperparameter for the decision tree is 2 and 10 for max depth and minimum samples leaf, respectively. Support vector machine runs optimally with hyperparameter values of 0.1, 1, and rbf for C, gamma, and kernel, respectively. Lastly, the random forest algorithm runs optimally with hyperparameter values of 5 for the number of estimators, auto for max features, 10 for max depth, 10 for minimum samples split, 4 for minimum samples leaf, and true for bootstrap.

Table. 5. The best hyperparameter value for each model.

Model	Hyperparameter	Value
Decision Tree	Max Depth	2
	Minimum Samples Leaf	10
Support Vector Machine	C	0.1
	Gamma	1
	Kernel	rbf
Random Forest	Number of Estimators	5
	Max Features	auto
	Max Depth	10
	Minimum Samples Split	10
	Minimum Samples Leaf	4
	Bootstrap	True

4.3 Model Evaluation

Along with the hyperparameter tuning process, we evaluated the performance results of each model. We used metrics like mean squared error (MSE) and the R2 value to determine the best model. In the end, we chose the best model for downscaling the TRMM satellite image.

Table. 6. The performance of each model.

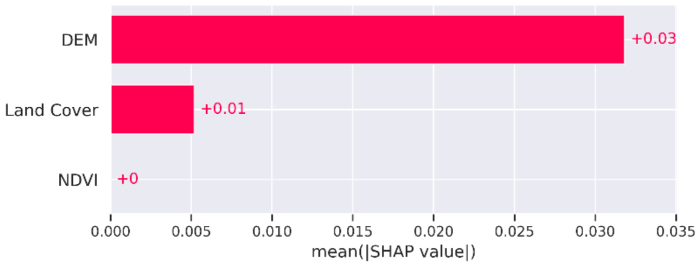
Model	MSE	R^2
Random Forest (RF)	0.00071	0.51952
Support Vector Machine (SVM)	0.00192	-0.30248
Decision Tree (DT)	0.00048	0.67107
Multiple Linear Regression (MLR)	0.00088	0.40327

Based on the table above, we observedthat the decision tree model had the best performance with its MSE value of 0.00048 and the R2 value of 0.67107. Following that model, the random forest had the second-best performance with its MSE value of 0.00071 and R2 value of 0.51952. The multiple linear regression had the third-best performance with its MSE value of 0.00088 and R2 value of 0.40327. The support vector machine had the worst result of all models. Besides that, the model also had an anomaly result, in which the R2 value is negative. Therefore, we selectthe decision tree model for downscaling the TRMM satellite image based on the performance result.

4.4. Model's Feature Importance

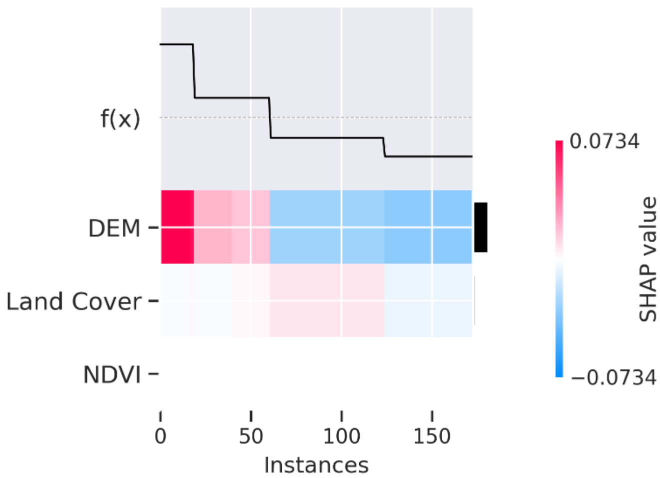
Besides modeling the data, we also checked the feature importance from the best model, which is the decision tree model. We did this to make sure the model could explain the reason behind each prediction. We used the SHAP library for visualizing the variable contributions to the prediction result.

Figure. 8. The variable importance from the model.



There are two visualizations we created. The first one is the feature importance bar chart. Based on the chart above, we saw that the DEM variable had the most significant contribution to the prediction result, where the variable had a 0.03 score of SHAP value. The land cover variable placed second with a SHAP value of 0.01. Interestingly, the NDVI was in the last position. It was because the NDVI variable had a SHAP value score of 0.

Figure. 9. Variable importance on each data cluster.

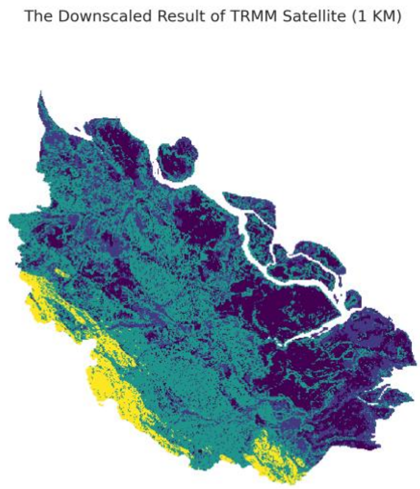


The second chart is the heat map chart. It describes the impact of each variable on the data that aredivided into several clusters. Based on the chart above, Weconcludethat if an observation contained a high value of the DEM variable, the precipitation value tends to be higher. As the DEM value became lower, the lower the precipitation value was. The land cover tends to work differently. If the DEM value was in the middle, like on the center cluster, the impact from the land cover variable tends to be higher and lower if the DEM variable gets a lower value. The SHAP value of NDVI tends to be neutral. Therefore, the NDVI didn't impact the prediction result.

4.5. Result Validation

After evaluating the model, we used it to downscale the TRMM satellite by taking the proposed inputs like NDVI, DEM, and Land Cover. By having the result, we sampled the data based on the coordinates of the existing rain gauge station in Riau Province, Indonesia, to gather the downscaled precipitation data. Based on our modeling result, we found an error between the real and downscaled precipitation value. Figure 10 shows the downscaling result from the random forest model.

Figure. 10. Result of downscaling process from the TRMM satellite image.



In Pekanbaru, the real precipitation value is 0.1 mm/year, while the downscaled result is 0.31386 mm/year. While in Kampar and Japura, both have the same actual precipitation value, which is 0 mm/year, and the same downscaled annual precipitation with the precipitation value is 0.31386 mm/year. Besides the modeling process, one of the reasons why the data have such an error margin is there are lots of missing values from all stations, whether the reason is that the station didn't measure all the times, which leads to bias on the result. Also, the number of rain gauge stations in Riau Province is not spatially distributed. Therefore, we cannot be certain of the evaluation result to determine whether the downscaling result is suitable to complement the ground station measurement. To mitigate this, we recommend enlarging the study area. Also, to look at how well the model performs, we recommend applying the downscaling process in a monthly term to see whether the model is robust to a monthly scale.

Table. 7. Actual and downscaled precipitation data.

Station's Name	Location	Actual Precipitation	Downscaled Precipitation
Stasiun Meteorologi Sultan Syarif Kasim II	Pekanbaru	0.1	0.31386
Stasiun Klimatologi Kampar	Kampar	0	0.31386
Stasiun Meteorologi Japura	Japura	0	0.31386

5. Conclusion

The study succeeds in applying the spatial downscaling process using machine learning models in Riau Province, Indonesia. We found the decision tree model has the best performance with the MSE value of 0.00048 and the R^2 value of 0.67107. We also found that the DEM variable is the most important variable on the downscaling process, followed by the NDVI and the land cover variable. Although there are errors in the validation result, the downscaled result has a prospect for complementing results with the rain gauge stations. For further research, we recommend expanding the study area because of the inadequate amount of rain gauge station and study about the effect of the downscaling model in a monthly term.

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., & Nelkin, E. (2003). The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present). *Journal of Hydrometeorology*, *4*(6), 1147–1167. https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2

Bell, J., Herring, S., Jantarasami, L., Adrianopoli, C., Benedict, K., Conlon, K., Escobar, V., Hess, J., Luvall, J., Pérez García-Pando, C., Quattrochi, D., Runkle, J., & Iii, S. (2016). *Ch. 4: Impacts of Extreme Events on Human Health* (pp. 99–128). https://doi.org/10.7930/J0BZ63ZV

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Damien, S.-M. (2019). *MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006* [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MCD12Q1.006

Elnashar, A., Zeng, H., Wu, B., Zhang, N., Tian, F., Zhang, M., Zhu, W., Yan, N., Chen, Z., Sun, Z., Wu, X., & Li, Y. (2020). Downscaling TRMM Monthly Precipitation Using Google Earth Engine and Google Cloud Computing. *Remote Sensing*, *12*(23), 3860. https://doi.org/10.3390/rs12233860

Gessesse, A. A., & Melesse, A. M. (2019). Chapter 8—Temporal relationships between time series CHIRPS-rainfall estimation and eMODIS-NDVI satellite images in Amhara Region, Ethiopia. In A. M. Melesse, W. Abtew, & G. Senay (Eds.), *Extreme Hydrology and Climate Variability* (pp. 81–92). Elsevier. https://doi.org/10.1016/B978-0-12-815998-9.00008-7

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

James, G., Witten, D., Hastie, T., & Tibshirani, R. (Eds.). (2013). *An introduction to statistical learning: With applications in R*. Springer.

Jarvis, A., Reuter, H., Nelson, A., & Guevara, E. (2008). Hole-filled SRTM for the globe version 3, from the CGIAR-CSI SRTM 90m database. *See Http://Srtm. Csi. Cgiar. Org*.

Jing, W., Yang, Y., Yue, X., & Zhao, X. (2016). A Spatial Downscaling Algorithm for Satellite-Based Precipitation over the Tibetan Plateau Based on NDVI, DEM, and Land Surface Temperature. *Remote Sensing*, *8*(8), 655. https://doi.org/10.3390/rs8080655

Kamel, D. (2015). *MOD13A2 MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid V006* [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD13A2.006

Lezama Valdes, L.-M., Katurji, M., & Meyer, H. (2021). A Machine Learning Based Downscaling Approach to Produce High Spatio-Temporal Resolution Land Surface Temperature of the Antarctic Dry Valleys from MODIS Data. *Remote Sensing*, *13*(22), 4673. https://doi.org/10.3390/rs13224673

Liu, Y., Yang, Y., Jing, W., & Yue, X. (2018). Comparison of Different Machine Learning Approaches for Monthly Satellite-Based Soil Moisture Downscaling over Northeast China. *Remote Sensing*, *10*(1), 31. https://doi.org/10.3390/rs10010031

Mahmud, Mohd., Hashim, M., Matsuyama, H., Numata, S., & Hosaka, T. (2018). Spatial Downscaling of Satellite Precipitation Data in Humid Tropics Using a Site-Specific Seasonal Coefficient. *Water*, *10*(4), 409. https://doi.org/10.3390/w10040409

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports*, *10*(1), 13768. https://doi.org/10.1038/s41598-020-70816-2

Xu, S., Zhao, Q., Yin, K., He, G., Zhang, Z., Wang, G., Wen, M., & Zhang, N. (2021). Spatial Downscaling of Land Surface Temperature Based on a Multi-Factor Geographically Weighted Machine Learning Model. *Remote Sensing*, *13*(6), 1186. https://doi.org/10.3390/rs13061186

Yan, R., & Bai, J. (2020). A New Approach for Soil Moisture Downscaling in the Presence of Seasonal Difference. *Remote Sensing*, *12*(17), 2818. https://doi.org/10.3390/rs12172818

Zhan, C., Han, J., Hu, S., Liu, L., & Dong, Y. (2018). Spatial Downscaling of GPM Annual and Monthly Precipitation Using Regression-Based Algorithms in a Mountainous Area. *Advances in Meteorology*, *2018*, 1506017. https://doi.org/10.1155/2018/1506017