Classification of Flood Alert in Jakarta with Random Forest

1st Grady, F.
Computer Science Department,
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
febryan.grady@binus.ac.id

2nd Tarigan, J. K.
Computer Science Department,
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
joelry.keegan@binus.ac.id

3rd Wahidiyat, J. R.
Computer Science Department,
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
joel.wahidiyat001@binus.ac.id

4th Prasetyo, A.
Computer Science Department,
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
anang.prasetyo@binus.ac.id

Abstract—One of the disasters currently considered the biggest in Indonesia is flooding which needs to be considered and will be discussed in the research in this paper. A flood is a natural event that occurs when spilt water reaches the mainland. The dangers caused by it are not small. Therefore, this paper aims to determine the classification of flood alerts in Jakarta using Random Forest algorithm. The datasets used in this study are from Open Data Jakarta and BMKG. We divide our flood alert status into four parts, namely 'normal', 'flood alert 1', 'flood alert 2', and 'flood alert 3'. For this study, we will be using the Random Forest algorithm. Random Forest is a machine learning algorithm that is usually used to classify large amounts of data. Due to random forest being a collection of tree classifiers, the number of tree classifiers must be determined in advance for modeling. In addition to collecting datasets, we also check and ensure that the datasets that have been collected meet the criteria we need. The next stage is that we will preprocess the data to fix incorrect or empty data so that the modeling process can give better results. As a result, we got 86% accurate in classifying and found several attributes that have major influence on flooding in Jakarta. The top three attributes are: 'floodgate_name', 'location', and 'month'. This means that flooding in Jakarta usually happens in certain times and places. These attributes can be used as a benchmark to be able to overcome future flooding events in Jakarta.

Index Terms—machine learning, flood, random forest, classification, prediction, jakarta

I. Introduction

Floods are one of the natural disasters that often occur in any part of the world, including Indonesia. Floods have greatly impacted the lives of Indonesian people, around 119 people in 2018 died from floods, 228,130 houses were flooded, and 738 public facilities were damaged by it [1]. According to the Indonesian dictionary, a flood is when the land (which is usually dry) sinks due to a rising water volume. The cause of the flooding is at least influenced by three important factors, such as dynamics and urban development, urban demography, and land use as well as land-use change [2].

Floods in Indonesia have occurred in various cities, especially Jakarta. Based on 'Badan Meteorologi, Klimatologi, dan Geofisika' (BMKG) data in 2020, Jakarta has the highest rainfall (377mm/day) compared to 154 years ago (185.1mm/day). Various efforts have been made to overcome floods in Jakarta, ranging from traditional efforts such as reforestation to modern using Jakarta Kini (JAKI) application. Quoted from the Jakarta Flood Monitoring website, JAKI is a flood monitoring applica- tion with features such as notification of flood points in Jakarta and reporting flood inundation. However, this application does not have a feature to predict floods. To be able to predict floods, a sophisticated technology known as Machine Learning needs to be used.

A computational algorithm where it only needs input data, will give the desired output without the need for programming, and improve itself without the help of a user is called Machine Learning. With Machine Learning we can finish tasks that were previously difficult to do, for example the capabilities to analyze the capital market, analyze the habits of smartphone users, analyze a photo, and many more [3].

Based on machine learning capabilities, in this research, we want to use machine learning to be able to classify flood alert status. Flood alerts are made so that before the occurrence of a flood, the flood alert can be informed by the officer and people living in the area can evacuate and be alert at once. Flood alert status in Indonesia is divided into 4 levels, namely 'normal', 'flood alert 1', 'flood alert 2', and 'flood alert 3'. 'normal' is a situation where the water alert status is still stable, and there is no clear possibility of flooding, 'flood alert 1' is a state where the water level must be watched out for if, for several hours, the water level has not receded or is in a dangerous condition, 'flood alert 2' is a condition where the water level causes inundation and spreads, 'flood alert 3' is a condition with a puddle of water in the area, but it is not critical or dangerous.

It is hoped that by conducting this research, disaster management in Jakarta can be carried out more accurately and community preparedness can be maximized so that the risk of major flooding can be carried out.

II. Literature Review

Flood is a word or term that is familiar to everyone. [4] According to (Paul Mun˜oz et al, 2018), a common natural disaster phenomenon that occurs throughout the world is flooding. It is expected that in the next few years, flood events will be more intensive due to an increase in extreme rainfall. Floods have many negative impacts on the natural environment. In catchment areas with highlands, flood events such as flash floods can cause fatal damage to downstream infrastructure and have a socio-economic impact. Emergency conditions are a threat to life, nature, and the ecosystem as a whole [5]. The direct effects of flooding include the loss of life, destruction of property, loss of crops and livestock, and worsening health conditions as a result of waterborne diseases. [6].

In order to overcome the adverse effects caused by this disaster, flood prediction is very necessary to prevent damage. However, flood prediction is one of the most challenging, difficult, and important problems in the field of hydrology due to its large contribution in reducing economic and life losses [7]. Various difficulties faced in flood forecasting are the need to identify hazards and assess the likelihood of floods, dependence on climate change measurements of rain, heat, and wind intensity, as well as predicting errors that can cause increased damage [8]. Due to climate change has increased the frequency and intensity of heavy rain events, frequently resulting in flooding damage in urban areas. Since the 1980s, rainfall events with intensities greater than 30 mm/hour have occurred more frequently. [9].

According to the analysis, the hydrological and environmen- tal predictors, literacy, land ownership, and house structures are important for prevention measures and disaster- related factors show less significance [10]. Therefore, identifying flood-prone areas is important to reduce the loss of life and property [11]. There are various ways that can be done to prevent flooding. One of which is the best and simplest early warning system is to use AI algorithms for flood forecasting due to heavy rains and flooding in various water bodies. With the sensor, we can obtain a wide range of datasets to create a framework in Machine Learning so that the predic- tion results obtained are strong, proficient, and precise [12]. Researchers have suggested various models to assess flood hazards. Hydrological models, hydrodynamic models, multi- criteria decision analysis (MCDA), statistical models (SM), and machine learning technologies (ML) that are integrated in geographic information systems (GIS) have been mainly the focus of most of the models [13]. Recently, machine learning models have been designed and used to handle high-dimensional illustrations by unifying supervised learning algorithms and feature selection [14].

One of the algorithms in Machine Learning is Random Forest (RF), the most established tree-based system, which is one of the algorithms in Machine Learning. Random Forest is a classification and regression tree to overcome the problem of overfitting the decision tree while maintaining predictive accuracy. This technique was developed by Breiman (2001), becoming a popular tool in the geosciences field due to its flexibility and availability in various software [15]. Using four indications (precision, recall, accuracy, and F-score) as well as a receiver operating characteristic (ROC) curve was adopted to evaluate the performance of the model in the training and test data set [16].

In research, various solutions have been found to realize flood forecasting. One of the solutions that can be used is LSTM. [17] A research study conducted by Vinayaka Gude, et al. found that deep learning can be used to predict floods. The Long Short-Term Memory (LSTM) algorithm, which is one type of architecture, can help realize flood forecasting. An iterative neural network with the ability to store long-term data is called an LSTM. In order to transform data, LSTM cells add or subtract information under the control of gates and add or multiply vectors. Another study conducted by Laura Lopez- Fuentes, et al. found that flood detection can also be done by comparing a collection of flood photos on social media. His research shows that incorporating images and additional information (metadata) will significantly improve system performance [18].

In their journal, Marcell Motta et al. developed a flood prediction system that combines Machine Learning classifiers with GIS techniques in order to be an effective tool for urban management and resilience planning Random Forest, K- Nearest Neighbors (KNN), and Multi-Layer Perceptron are employed (MLP). Despite relying solely on weather conditions and a lack of spatial variability for certain conditions, this approach can consistently predict floods with a high level of confidence. [19]. Research to detect damage using machine learning is explained in [20] and the results obtained are from several models used by researchers, namely CRISP-DM, BN, DT, KNN, and SVM, it was found that the BN model has slightly better accuracy than other models. Research conducted by Khabat Khosravi, et al. regarding flooding located in the southeastern part of Jiangxi Province (length 115°), Ningdi County, one of the most flood-prone areas in China, was selected as a field of study using the NBT algorithm with valid validation. Resulted in the best performing models (NBT and NB) (AUC = 0.98), indicating that, for the models that are studied, the NBT model is regarded as a promising tool for assessing flood-prone areas, allowing for appropriate flood hazard planning and management. [21].

Experiments on the application of Machine Learning have been carried out in various locations. For example, the Machine Learning model in the form of Random Forest has been applied in the United States by Zahura, F. T., et al. [22]. In addition, there is also the Ensemble Hybrid model, which was applied in Bangladesh by Towfiqul Islam, A. R. M., et al. [23]. Machine Learning can improve the accuracy of flood predictions. This has been investigated by several people, such as researcher Ke, Q. et al., who were testing the Machine Learning model to predict flood events in Shenzhen, China, found that the Machine Learning model has higher accuracy than conventional empirical methods. [24]. In addition, there is also a study by Esfandiari, M., et al., which found that the Ma- chine Learning Pseudo-Random Forest model has higher ac- curacy than the conventional method of Height Above Nearest Drainage (HAND) [25]. However, before making a machine learning model to predict floods, there is an important thing to note. Algorithms from machine learning will determine the quality of machine learning that will be made. The more suitable the algorithm used, the better the prediction given. This will be very influential when implementing machine learning that has been made public.

III. PROPOSE METHOD

A. Random Forest

Random forest, a machine algorithm developed by Breiman in 2001 [26], is a combination of tree classifiers where each tree will vote for the most popular class from a dataset. After voting, the voting will be combined to get the result. Because a random forest is a collection of tree classifiers, the number of tree classifiers must be determined in advance for modeling. The machine learning method we chose is Random Forest because the random forest is not only easy to use but has recognizable accuracy and the ability to handle several types of samples and features with high spatial dimensions. Random Forest has a high potential to handle multiple kinds of problems, such as cases in real-life systems, in our case on flood prediction. Based on the paper we found [27], there is an opinion that the greater number of trees specified does not always increase the performance of random forest more significantly because with a small number of trees and with a large number of trees, the given performance may give the same results. An overview of the random forest process can be seen in Fig. 1.

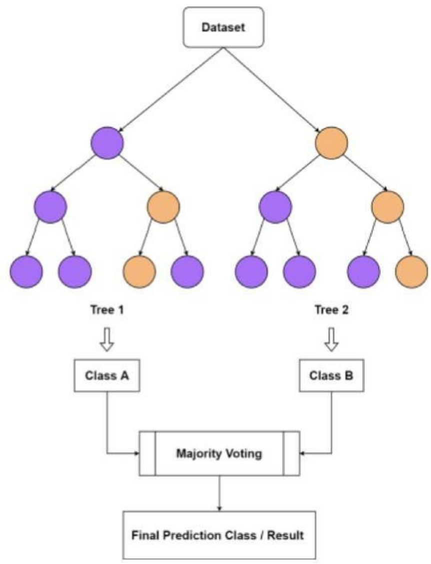

Fig. 1. Random Forest Process

B. Research Design

The steps for conducting our research are as shown in Fig. 2. At the initial stage, we will collect datasets related to our research topic. In addition to collecting datasets, we also check and ensure that the datasets that have been collected meet the criteria we need. The next stage is that we will preprocess the data to fix incorrect or empty data so that the modeling process can give better results. We will also analyze the data based on the results of the model that has been made and draw conclusions based on the results of the research data analysis. Based on the data that we analyzed, we will tune the model to become more optimal and efficient.

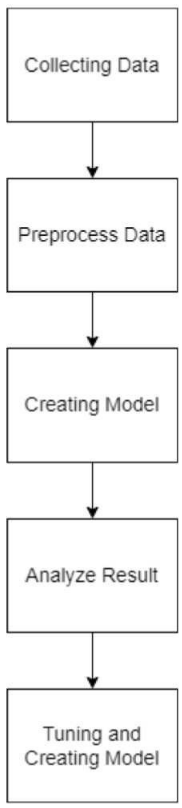

Fig. 2. Research Design

C. Datasets

To be able to classify flood alerts in Jakarta, we collected datasets related to water level data in DKI Jakarta province from 2019 - 2021. The datasets were obtained from the Jakarta

TABLE I
JAKARTA OPENDATA ATTRIBUTE LISTS

No	Attribute	Details
1	floodgate_name	Location to make observations
2	location	Observable objects (sea, river or river name, etc.)
3	lattitude	Latitude of the location
4	longitude	Longitude of the location
5	date	Date and time the data was taken
6	water_level	Water height (centimeter)
7	alert_status	Alert status for water level

OpenData website [28], which has the following attribute sets on Table I.

The result of only using ‘floodgate_name’ and ‘water_level’ attributes is less significant for our research even though it is sufficient to predict ‘alert_status’. This is because most floodgates have an indicator of alert status based on the amount of water in their area [29]. The following Fig. 3. is an example of a ‘alert status’ indicator on one of the floodgates.

Fig. 3. Floodgates indicator of alert status

Therefore, for our research to provide more significant results, we collected additional datasets from the BMKG Online Data site [30]. This additional dataset has the following attribute sets on Table II.

The Jakarta OpenData and BMKG datasets will be combined by using the ‘date’ attribute from each dataset so that data analysis becomes more manageable with a comprehensive approach and can also consider the information available based on the data obtained. In the decision-making process, it will be better because the amount of data collected makes the level of decisions taken can be wrong reduced.

D. Preprocessing

In the dataset, we found that the dataset has 10 quantitative attributes, such as ‘Tn’, ‘Tx’, ‘Tavg’, ‘RH_avg’, ‘RR’, ‘ss’, ‘ff_x’, ‘ddd_x’, ‘ff_avg’, and ‘water_level’. We also found that the dataset has 4 categorical attributes, such as

TABLE II
BMKG DATASET ATRRIBUTE LIST

No	Attribute	Details
1	date	Date and time the data was taken
2	Tn	Minimum temperature
3	Tx	Maximum temperature
4	Tavg	Average temperature
5	RH_avg	Average humidity
6	RR	Rainfall
7	ss	The duration of the sun
8	ff_x	Maximum wind speed
9	ff_avg	Average wind speed
10	ddd_x	Wind direction at maximum speed
11	ddd_car	Most wind direction

‘floodgate_name’, ‘location’, ‘date’, and ‘ddd_car’. From the data that we have collected, to make the data more reliable and better for Machine Learning, we need to clean up the data before creating the model. In this preprocessing part, we remove the ‘water_level’ attribute as it is not required in this research, we also change all the empty data to be the mean of the data for quantitative data and mode of the data for qualitative data. We also combine the flood data and weather data based on the date or ‘Date’ attribute of the data. For the last part, we also split the date of the data to be month only and day to analyze whether day and month is also contributing to flooding in Jakarta.

We found the dataset that we have collected is not well balanced, as the ‘Status: Normal’ class has more data points compared to other classes. To be more specific, the count data of each class is 570441 for ‘Status: Normal’, 64508 for ‘Status: Flood Alert 3’, 36208 for ‘Status: Flood Alert 2’, and 2085 for ‘Status: Flood Alert 1’. To mitigate this problem, we are using the naive RandomOverSampler from sklearn to fill up the gap between each class. We use the naive method to make sure the data is more accurate to the real situation. We also maximize the maximum number of each Alert class to be 1:3 from Normal class. The maximization is to make sure the model does not bias toward Alert than Normal as more Normal data are more accurate in the real situation.

E. Tuning Model

In Random Forest, to make the model faster to be built and not waste any computational power, we need to tune the model. For the tuning part, we modify how many trees are in the model, and what is the maximum number of depths in each tree. With those parameters being tuned, we can achieve better computational speed and power in the tuned model.

IV. EXPERIMENTAL RESULTS

For both models, the untuned model with 100 trees and 32 average depths and tuned model with only 10 trees and 20 maximum depths, we get the accuracy represented in Table III.

Based on our experiment that we have done, we found that when tuning is done, the model creation is much faster. On our case with specification of AMD Ryzen 7 4700H processor

TABLE III
ACCURACY OF UNTUNED AND TUNED MODEL

	Precision	Recall	F1-Score	Support
Normal	0.97	0.93	0.95	113901
Flood Alert 3	0.89	0.99	0.93	38188
Flood Alert 2	0.66	0.76	0.71	38168
Flood Alert 1	0.73	0.63	0.67	37920
Accuracy			0.86	228177
Macro avg	0.81	0.83	0.82	228177
Weighted avg	0.87	0.86	0.86	228177

and 16GB of RAM we got a comparison result of untuned and tuned model shown in Table IV.

TABLE IV
DIFFERENCE OF MODEL CREATION TIME

	Untuned Model	Tuned Model
Parameters	100 estimators and 32.3 depths	10 estimators and 20 depths
Model Creation Time	137.6 seconds	16.2 seconds

As shown on Table IV, we can conclude that the tuned model is approximately 10x faster than the untuned model. With the model that we created we also found out which independent variables affect the categorizing of floods in Jakarta. Those variables can be analyzed with the Random Forest's feature importance. The feature importance is calculated based on the Random Forest model that we have created. By using scikit-learn Random Forest's feature_importances_ attribute, we can show all the features that shape how the Random Forest is created. The feature importance list can be seen in Table V.

TABLE V
DIFFERENCE OF MODEL CREATION TIME

Variables	Importance Level
floodgate name	0.384
location	0.203
month	0.063
RR	0.061
day	0.049
ss	0.046
Tavg	0.044
Tx	0.043
ddd x	0.035
Tn	0.032
ff x	0.021
fff avg	0.008
ddd car	0.005

As shown on Table V, we found that the top 3 variables that easily categorize floods in Jakarta are 'floodgate_name', 'location', and 'month'. This means that floods in Jakarta usually happen in only certain places and times. The flood in Jakarta can also be analyzed based on the weather, as we can see that the RR variable is the number 4 affecting variable in this case.

V. CONCLUSIONS

In this paper, we conduct a study to classify flood alerts in Jakarta using the Random Forest model. The dataset we use is a combination of datasets from Jakarta OpenData and BMKG. We combine the datasets so that the model can provide more significant flood preparedness prediction results. We also tune each class to get the proper comparison so that the results of the predictions cannot be adjusted to the actual situation of the prediction experiment with a ratio of 1:3 from each class to the 'normal' class.

Based on the experiments carried out, we obtained 86accuracy of the data, and divided the alert status into four groups. We also found several attributes that have major influence on flooding in Jakarta. These attributes can be used as a benchmark to be able to overcome future flooding events in Jakarta.

REFERENCES

[1] Rosyida, A., & Nurmasari, R. (2019). Analisis Perbandingan Dampak Kejadian Bencana Hidrometeorologi dan Geologi di Indonesia Dilihat Dari Jumlah Korban (Studi: Data Kejadian Bencana Indonesia 2018). Jurnal Dialog dan Penanggulangan Bencana, 10(1), 12-21.

[2] Eldi, E. (2020). ANALISIS PENYEBAB BANJIR DI DKI JAKARTA. Jurnal Inovasi Penelitian, 1(6), 1057-1064.

[3] El Naqa, I., & Murphy, M. J. (2015). What Is Machine Learning? In Machine Learning in Radiation Oncology (pp. 3±11). Springer International Publishing. https://doi.org/10.1007/978-3-319-18305-3 1

[4] Mun˜oz, P., Orellana-Alvear, J., Willems, P., & CeÁlleri, R. (20T8). Flash- Flood Forecasting in an Andean Mountain Catchment—Development of a Step-Wise Methodology Based on the Random Forest Al-
gorithm. In Water (Vol. 10, Issue 11, p. 1519). MDPI AG. https://doi.org/10.3390/w10111519

[5] Inyang, U. G., Akpan, E. E., & Akinyokun, O. C. (2020). A Hybrid Machine Learning Approach for Flood Risk Assessment and Classification. In International Journal of Computational Intelligence and Applications (Vol. 19, Issue 02, p. 2050012). World Scientific Pub Co Pte Lt. https://doi.org/10.1142/s1469026820500121.

[6] Sankaranarayanan, Suresh, Malavika Prabhakar, Sreesta Satish, Prerna Jain, Anjali Ramprasad, and Aiswarya Krishnan. "Flood prediction based on weather parameters using deep learning." Journal of Water and Climate Change 11, no. 4 (2020): 1766-1783.

[7] Kinage, C., Kalgutkar, A., Parab, A., Mandora, S., & Sahu, S. (2019). Performance Evaluation of Different Machine Learning Based Algorithms for Flood Prediction and Model for Real Time Flood Prediction. In 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA). 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA). IEEE. https://doi.org/10.1109/iccubea47591.2019.9128379

[8] Idan, H. M., & Hussein, K. Q. (2021). Comparison study between selected techniques of (ML, SVM and Deep Learning) regarding prediction of Flooding in Eastof Iraq. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(14), 2893-2904.

[9] Keum, H. J., Han, K. Y., & Kim, H. I. (2020). Real-Time Flood Disaster Prediction System by Applying Machine Learning Technique. In KSCE Journal of Civil Engineering (Vol. 24, Issue 9, pp. 2835±2848). Springer Science and Business Media LLC. https://doi.org/10.1007/s12205-020-1677-7

[10] Schoppa, L., Disse, M., & Bachmair, S. (2020). Evaluating the performance of random forest for large-scale flood discharge simulation. In Journal of Hydrology (Vol. 590, p. 125531). Elsevier BV. https://doi.org/10.1016/j.jhydrol.2020.125531

[11] Rahman, M., Ningsheng, C., Islam, M. M., Dewan, A., Iqbal, J., Washakh, R. M. A., & Shufeng, T. (2019). Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis. In Earth Systems and Environment (Vol. 3, Issue 3, pp. 585±601). Springer Science and Business Media LLC. https://doi.org/10.1007/s41748- 019-00123-y

[12] K. Kunverji, K. Shah, and N. Shah, ªA Flood Prediction System Developed Using Various Machine Learning Algorithms," SSRN Electronic [Online]. Available: Journal. Elsevier 2021 BV, http://dx.doi.org/10.2139/ssrn.3866524

[13] Rahman, M., Ningsheng, C., Islam, M. M., Dewan, A., Iqbal, J., Washakh, R. M. A., & Shufeng, T. (2019). Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis. In Earth Systems and Environment (Vol. 3, Issue 3, pp. 585±601). Springer Science and Business Media LLC. https://doi.org/10.1007/s41748- 019-00123-y

[14] Khalaf, M., Hussain, A. J., Al-Jumeily, D., Baker, T., Keight, R., Lisboa, P., Fergus, P., & Al Kafri, A. S. (2018). A Data Science Methodology Based on Machine Learning Algorithms for Flood Severity Prediction. In 2018 IEEE Congress on Evolutionary Computation (CEC). 2018 IEEE Congress on Evolutionary Computation (CEC). IEEE. https://doi.org/10.1109/cec.2018.8477904

[15] Schoppa, L., Disse, M., & Bachmair, S. (2020). Evaluating the performance of random forest for large-scale flood discharge simulation. In Journal of Hydrology (Vol. 590, p. 125531). Elsevier BV. https://doi.org/10.1016/j.jhydrol.2020.125531

[16] Zhao, G., Pang, B., Xu, Z., Peng, D., & Xu, L. (2019). Assessment of urban flood susceptibility using semi-supervised machine learning model. In Science of The Total Environment (Vol. 659, pp. 940±949). Elsevier BV. https://doi.org/10.1016/j.scitotenv.2018.12.217

[17] Gude, V., Corns, S., & Long, S. (2020). Flood Prediction and Uncer- tainty Estimation Using Deep Learning. In Water (Vol. 12, Issue 3, p. 884). MDPI AG. https://doi.org/10.3390/w12030884

[18] Lopez-Fuentes, L., Weijer, J.V., Bolan˜os, M., & Skinnemoen, H. (2017). Multi-modal Deep Learning Approach for Flood Detection. MediaEval.

[19] Motta, M., de Castro Neto, M., & Sarmento, P. (2021). A mixed approach for urban flood prediction using Machine Learning and GIS. In International Journal of Disaster Risk Reduction (Vol. 56, p. 102154). Elsevier BV. https://doi.org/10.1016/j.ijdrr.2021.102154

[20] Razali, N., Ismail, S., & Mustapha, A. (2020). Machine learning ap- proach for flood risks prediction. IAES International Journal of Artificial Intelligence, 9(1), 73.

[21] Khosravi, K., Shahabi, H., Pham, B. T., Adamowski, J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H.-B., GroÅf, G., Ho, H. L., Hong, H., Chapi, K., & Prakash, I. (2019). A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. In Journal of Hydrology (Vol. 573, pp. 311±323). Elsevier BV. https://doi.org/10.1016/j.jhydrol.2019.03.073

[22] Zahura, F. T., Goodall, J. L., Sadler, J. M., Shen, Y., Morsy, M. M., & Behl, M. (2020). Training Machine Learning Surrogate Models From a High -Fidelity Physics -Based Model: Application for Real -Time Street-Scale Flood Prediction in an Urban Coastal Community. In Water Resources Research (Vol. 56, Issue 10). American Geophysical Union (AGU). https://doi.org/10.1029/2019wr027038

[23] Towfiqul Islam, A. R. M., Talukdar, S., Mahato, S., Kundu, S., Eibek, K. U., Pham, Q. B., Kuriqi, A., & Linh, N. T. T. (2021). Flood susceptibility modelling using advanced ensemble machine learning models. In Geoscience Frontiers (Vol. 12, Issue 3, p. 101075). Elsevier BV. https://doi.org/10.1016/j.gsf.2020.09.006

[24] Ke, Q., Tian, X., Bricker, J., Tian, Z., Guan, G., Cai, H., Huang, X., Yang, H., & Liu, J. (2020). Urban pluvial flooding prediction by machine learning approaches ± a case study of Shenzhen city, China. In Advances in 103719). Elsevier (Vol. 145, Water BV. Resources p. https://doi.org/10.1016/j.advwatres.2020.103719

[25] Esfandiari, M., Abdi, G., Jabari, S., McGrath, H., & Coleman, D. (2020). Flood Hazard Risk Mapping Using a Pseudo Supervised Random Forest. In Remote Sensing (Vol. 12, Issue 19, p. 3206). MDPI AG. https://doi.org/10.3390/rs12193206 -

[26] Biau, G., & Scornet, E. (2016). A random forest guided tour. In TEST (Vol. 25, Issue 2, pp. 197±227). Springer Science and Business Media LLC. https://doi.org/10.1007/s11749-016-0481-7

[27] Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How Many Trees in a Random Forest? In Machine Learning and Data Mining in Pattern 154±168). Springer Berlin Heidelberg. (pp. Recognition https://doi.org/10.1007/978-3-642-31537-4_13

[28] ªKumpulan Dataset: Open Data Jakarta." Jakarta Open Data, https://data.jakarta.go.id/dataset?q=tinggi+muka+air

[29] Tinggi Muka Air Online — Dinas Sumber Daya Air Provinsi DKI Jakarta, http://poskobanjirdsda.jakarta.go.id/Default.aspx

[30] Data Online - Pusat Database https://dataonline.bmkg.go.id/home