Flood Risk Mapping of Jakarta Using Genetic Algorithm Rule-Set Production (GARP) and Quick Unbiased Efficient Statistical Tree (QUEST) Methods

R R Yusya1*, M R Septyandy1, and T L Indra2

1Program Study Geology, Faculty of Mathematics and Natural Sciences, Universitas Indonesia
2Department of Geography, Faculty of Mathematics and Natural Sciences, Universitas Indonesia

*Email: tito.latif@sci.ui.ac.id

Abstract. Jakarta as the country’s capital, has experienced a series of floods that crippled cities in 2002, 2007 and 2013 and extreme rainfall has always been the main cause of major flood with casualties and property. Given the increasing impact of flooding in Jakarta, methods for assessing current and future flood risks is needed. A model is needed that can predict and determine which areas have the potential to be affected by future floods. Several models have been carried out for prediction and assessment of flood risk in Jakarta and several other major cities, namely GIS based on SVM models, hierarchical process analysis models (AHP), machine learning with ensemble models, and multivariate discriminant, classification, and regression trees. However, from several methods there are still some weaknesses, one of which is the time to get a predictive model that is quite long, which is 3-4 hours depending on the amount of data. To overcome this problem, genetic algorithm rule-set production (GARP) and quick unbiased efficient statistical tree (QUEST) models can create a prediction model with a shorter time with the same results. Several factors that influence flooding are used as input model data: precipitation, slope, distance to river, distance to channel, depth to groundwater, land use, elevation, and flood data for 2002 - 2019. The Area under receiver-operator characteristic curve (AUC-ROC) and root mean square error (RMSE) were used as evaluations for model performance. The results showed that the GARP model (AUC-ROC = 94%, RMSE = 0.2) had higher performance accuracy than the QUEST model (AUC-ROC = 84%, RMSE = 0.4). The results also indicated that the distance to channel, land use, and elevation as important parameters in determining flood hazards.

1. Introduction
Flood is one of the natural disasters of hydrometeorology that often occurs in Indonesia [1]. Flood disasters are caused by natural factors such as geographical and topographic conditions of the region, rainfall intensity, and the shape of the river channel [1]. Social factors are also related to flooding, such as population density and land use. The area of Indonesia that often experiences floods is the Special Capital Province of Jakarta (DKI Jakarta). Jakarta experienced several major flood disasters in 2002, 2007 and 2013 [2]. Floods that occur in Jakarta are caused by overflowing rivers that occur when rainfall is very high in the rainy season. The river overflowed due to it unable to accommodate the intensity of the rain that came and the silting of the river [3]. Impacts that occur due to floods are the paralysis of life activities in Jakarta and the impact of material and non-material losses such as the emergence of disease / health outbreaks, damage to buildings, dwellings, public facilities, and infrastructure, and can
cripple the economic activities of the local community. Therefore, proper risk management needs to be done so that the disaster can be reduced.

Modeling flood risk areas is a form of effort in predicting the occurrence of floods in the future. Flood risk modeling has been carried out with several GIS-based models such as the SVM (Support Vector Machine) model [4], the hierarchical analysis process model (AHP) [5], statistical models and machine learning [6], and multivariate discriminant, classification, and regression trees [7]. However, from some of these models there are some shortages, one of which is the time to get a fairly long prediction model that is around 3-4 hours depending on the amount of data used. To overcome these problems, the GARP and QUEST methods are able to create a prediction model with a shorter time with the same results [8].

Modeling using the GARP and QUEST methods has been used in predicting flood disasters in the City of Sarri, Iran [8]. Based on this research the writer wants to do research in the Jakarta area. The purpose of this study is to evaluate the factors that influence the occurrence of floods in Jakarta which will be included in the depth of the flood vulnerability map. The factors included in this model consist of seven components, namely: rainfall, land use, area height, slope, distance from large rivers, distance to river, and depth of ground water level.

2. Study Area and Input Data

2.1. Study Area

The Special Capital Region of Jakarta (5º 19 '12 " - 6º 23" 54 "S, 106º 22" 42 " - 106º 58" 18 "E) is the country's capital and largest city in Indonesia. Jakarta is located on the northwest coast of Java. Jakarta has a land area of about 661.52 km², with a population of 10,374,235 people (2017). The Jakarta area is the most populous metropolitan area in Southeast Asia and the second most populous city in the world. Jakarta was located in the lowlands at an average altitude of 8 m above sea level. The southern area of Jakarta is a mountainous area with high rainfall intensity. Jakarta is crossed by 13 large rivers which all lead to Jakarta Bay, namely: Ciliwung River, Pesanggrahan River, Angke River, Baru Barat River, Baru Timur River, Grogol River, Mookervart River, Krukut River, Cipinang River, Sunter River, Buaran River, Jati Kramat River, and Cakung River. In addition, there are two Flood Channels, namely the West Flood Canal and the East Flood Canal and two large drainages in the west and east of Jakarta, namely Cengkareng Drain and Cakung Drain. Ciliwung River is the largest river in Jakarta, located right in the middle of the city of Jakarta. Almost all of Jakarta's land areas are residential buildings or office buildings. Jakarta has hot, dry or tropical temperatures. Jakarta experiences the peak of the rainy season in January and February with an average rainfall of 350 millimeters with an average temperature of 27 °C. Rainfall between January and early February is very high, that's when Jakarta is flooded every year, and the peak of the dry season in August with an average rainfall of 60 millimeters. September and early October are very hot days in Jakarta, temperatures can reach 40 °C [9].

Figure 1. Administrative map of Jakarta city
2.2. Geological Study Area

In the geological map of Jakarta and the Thousand Islands [10], Jakarta is composed of the youngest to the oldest lithology: alluvium, coastal dike sediments, and alluvial fans. Alluvium in the form of clay to lump; alluvium fan in the form of fine tuff and sandy tuff intermittent tuff conglomerate; coastal embankment deposition of fine to coarse sand with mollusk shells (Figure 2).

![Regional geological map of Jakarta](image)

**Figure 2.** Regional geological map of Jakarta [10]

2.3. Data

The data used in this study amounted to 8 data consisting of 7 hydrological data which are factors that cause flooding, namely: rainfall data, land use data, area height data, slope percentage data, large river distance data, small river distance data, and groundwater level depth data [8] which are input data. As well as flood distribution data in Jakarta as external data. The input data used is shown in Table 1.

| No. | Data                        | Source                          |
|-----|-----------------------------|---------------------------------|
| 1   | Rainfall                    | BMKG                            |
| 2   | Land Use                    | USGS Landsat Imagery            |
| 3   | Elevation                   | DEM SRTM BIG                    |
| 4   | Slope                       | DEM SRTM BIG                    |
| 5   | Depth to Groundwater        | BPBD DKI Jakarta                |
| 6   | Distance to River           | BPBD DKI Jakarta                |
| 7   | Distance to Channel         | BPBD DKI Jakarta                |
2.3.1. Rainfall. The data used are rainfall data obtained from the recording of 5 observation station points belonging to the Meteorology, Climatology and Geophysics Agency (BMKG). The 5 station points are in the region:
- Kedoya Selatan, West Jakarta
- Pakubuwono, South Jakarta
- Pintu Air Pulo Gadung, East Jakarta
- Pasar Minggu, South Jakarta
- Teluk Gong, North Jakarta

Five data points are processed using ArcGIS software to get the value of interpolation, so that the entire Jakarta area has a different rainfall value. The interpolation results using the kriging method in ArcGIS produce a map of the rainfall distribution in Jakarta which is shown in Figure 3. The map shows that rainfall with high intensity is in the southeast of Jakarta while low intensity rain is in the northwest of Jakarta.

![Figure 3. Rainfall distribution map of Jakarta](image)

2.3.2. Land Use. Data used for land use were taken from USGS Landsat imagery which were then classified into four classes: built up land (settlements, buildings), water bodies (rivers, lakes and swamps), rice fields and fields, and vegetation cover. It can be seen in Figure 4 that land cover in the form of built up land is very dominant in the Jakarta area. The land cover of water bodies is mostly located in the northern part of Jakarta. Vegetation land cover and rice fields and fields only appear to be minimal.

![Figure 4. Land use map of Jakarta](image)
2.3.3. Elevation. Elevation data in the Jakarta area is obtained using digital elevation model data released by the Geospatial Information Agency (BIG). The DEM data is then processed using the ArcGIS application to classify the height data in the Jakarta area. Then obtained four height classes seen in Table 2.

| Class | Elevation |
|-------|-----------|
| 1     | <5 m      |
| 2     | 5-25 m    |
| 3     | 25-50 m   |
| 4     | >50 m     |

Elevation <5 m in the northern part of Jakarta including North Jakarta, part of West Jakarta and part of Central Jakarta. Elevation of 5-25 m is in the central part of Jakarta, covering parts of West Jakarta, Central Jakarta, South Jakarta and East Jakarta. Elevation of 25-50 m is in the southernmost part of Jakarta in the South Jakarta and East Jakarta regions. Elevation > 50 m is in the southern part of Jakarta bordering the Bogor area (Figure 5).

![Elevation Map of Jakarta](image)

2.3.4. Slope. The slope data in the Jakarta area is obtained using digital elevation model data issued by the Geospatial Information Agency (BIG). The DEM data is then processed using the ArcGIS application to classify the height data in the Jakarta area. Then obtained five slope classes with percentage seen in Table 3.

| Class | Percentage of Slope |
|-------|---------------------|
| 1     | 0 – 2 %             |
| 2     | 3 - 7 %             |
| 3     | 8 – 13 %            |
| 4     | 14 – 20 %           |
| 5     | >20 %               |
In contrast to the elevation data, the percentage of slope in the Jakarta area is spread randomly with the dominant percentage at 3-7%, while the slope of > 20% is also dominant around the watershed in Jakarta (Figure 6).

2.3.5. Depth to Groundwater. The depth of groundwater level data was obtained from BPBD DKI Jakarta. The data are three data points:
- Cilincing District, North Jakarta
- Pademangan District, North Jakarta
- Cakung District, East Jakarta

Just like rainfall data, groundwater depth data are interpolated to get data distribution that interprets groundwater depths in all points of DKI Jakarta province. Interpolation using the kriging method. The results can be seen in Figure 7 which shows that in the western part of Jakarta the depth of the groundwater level is getting deeper, while the eastern part of Jakarta is deeper and shallower.

2.3.6. Distance to River. Major rivers in the Jakarta area are 13 main rivers that flow in Jakarta, namely: Ciliwung River, Pesanggrahan River, Angke River, Baru Barat River, Baru Timur River, Grogol River, Mookervart River, Krukut River, Cipinang River, Sunter River, Buaran River, Jati Kramat River, and Cakung River. Including two flood canals, namely: West Flood Canal and East Flood Canal, and two large drainages in the west and east of Jakarta, namely: Cengkareng Drain and Cakung Drain.
To find out the distance from each river using the method in ArcGIS software, namely: Multiple Ring Buffer. Before using this method, first classifying river distance class which is divided into three namely near, intermediate and far as shown in table 4.

**Table 4. Distance to river classification**

| Class   | Distance       |
|---------|----------------|
| Near    | < 500 m        |
| Intermediate | 500 – 2500 m   |
| Far     | 2500 – 7000 m  |

The results of this method are seen in Figure 8. From that map, we can be seen that most of the areas in Jakarta are located at a medium distance from a large nearby river.

![Figure 8. Distance to river map of Jakarta](image)

**2.3.7. Distance to Channel.** The channel examined in this study is a tributary of the 13 major rivers in the Jakarta area. Class classification on distance data is the same as the distance on a large river but different on the size of the distance in each class is seen in Table 5.

**Table 5. Distance to channel classification**

| Class      | Distance       |
|------------|----------------|
| Near       | < 500 m        |
| Intermediate | 500 – 2500 m   |
| Far        | 2500 – 15000 m |
2.3.8. Flood Distribution. The data of flood point distribution was obtained from the Regional Disaster Management Agency (BPBD) of DKI Jakarta Province in 2015. It was seen that the distribution of flood points was most numerous in the northern area of Jakarta and along the main river flow (Figure 10).

3. Methods

3.1. Genetic Algorithm for Rule-Set Prediction (GARP)
Genetic Algorithm Rule-Set Production (GARP) is a machine learning algorithm that has demonstrated excellent predictive abilities in areas such as ecological modeling [11]. The GARP algorithm [8] produces predictions of flood inundation for urban areas through an iterative process to improve the stability of the model output. Running multiple processes to obtain different model outputs and using the best subset method is important in choosing the best output with optimal parameters. The set of models that achieve harmony between the limits of error of omission (sensitivity) and commission (specificity) are defined by the user [8]. GARP output is a collection of grids from the study area, which can be used in a GIS environment to identify flood-prone areas [8]. In this case, the GARP model is run using software such as Arc GIS. In addition, the importance of flood-causing factors such as rainfall, slope, distance to river, distance to channel, depth to ground water, land use, and elevation for urban flood hazard was analyzed using the GARP model.
3.2. Quick, Unbiased, and Efficient Statistical Tree (QUEST)

QUEST \cite{12} is a popular data processing model that produces subsets of data that are as homogenic as possible with respect to response variables. QUEST is a tree structured classification algorithm that produces a growing binary split decision tree \cite{8}. It uses a sequential tree planting method, which uses a linear discrimination analysis method in splitting tree nodes. This has many advantages over recursive tree construction methods such as classification and regression trees (CART) \cite{8}. The QUEST algorithm was chosen as the second model to predict flood inundation. In addition, the QUEST algorithm uses imputation instead of substitute separations to handle missing values. QUEST has negligible bias because it uses variable selection techniques that are unbiased in modelling. Therefore, QUEST can easily handle categorical and sustainable factors \cite{8}.

4. Result and Discussion

The results of the GARP model (AUC-ROC = 94%, RMSE = 0.2) show that this model uses a set of rules that have been restricted so that weights can be determined from each input data parameter used. From this model it is known that the main factors causing flooding from the weighting value of each parameter. The parameters of distance to rivers have the highest weighting of 30%, followed by height parameters of 20%, land use by 15%, distance to channels with weights of 15%, rainfall has a weight of 10%, ground water depth of 8%, and slope of 2%.

![Figure 11](image1.png)

**Figure 11.** 2015 flood distribution map results of the GARP Model

![Figure 12](image2.png)

**Figure 12.** 2019 flood prediction map results of the GARP Model
The results of the QUEST model (AUC-ROC = 84%, RMSE = 0.4) show that this model gets worse grades than the GARP model. This is because there is a need for data clustering that can produce bias values against the input data parameters used. The resulting weight values for each parameter are: height of 25%, land use by 20%, distance to rivers by 20%, distance to channels by 15%, rainfall by 10%, depth of ground water by 9%, and slope of 1%.

The weighting value of each of these parameters results in a disaster risk map that is classified into three classes, namely: small disaster risk, medium disaster risk, and large disaster risk. The map results in Figure 11 are the results of the GARP model for the 2015 flood distribution data and Figure 12 results from the GARP model for the 2019 flood prediction.

The results of the QUEST model can be seen in Figure 13 is the result of the QUEST model for 2015 flood distribution data and Figure 14 of the QUEST model results for the 2019 flood prediction.

The results of modelling using the GARP and QUEST models show that there is a similarity between the results of the model for the 2015 flood data distribution with the 2015 flood data distribution owned by DKI Jakarta BPBD. This can be seen from the distribution of flood data, which is mostly located in the northern area of Jakarta, very similar to the map of the model results which shows that the greatest risk level is in the northern area of Jakarta.

The results of the model for the predictions of floods in 2019 show that the area that experiences the greatest risk level becomes wider than the 2015 data. This is expected due to higher rainfall data and has a broad rainfall distribution. In the southern part of Jakarta, the point of flood risk is in the smallest risk class. In the central part of Jakarta, most are in the middle risk flood class.
5. Conclusion
Province of Special Capital Region of Jakarta is an area that is at risk for flood disasters. Then effective flood management is needed in predicting floods that will occur in the future. In this research, flood modelling was made using two models, namely the GARP model and the QUEST model. The GARP model and the QUEST model are used for research that produces flood risk maps in the Jakarta city area. The results of this model can be applied to flood zoning in urban areas. The results show the distance to the river is the main factor causing floods in the city of Jakarta. From these results it can be used using the GARP model and the QUEST model can be relied on with little data and is also fast in predicting flooding.

Acknowledgement
The Authors would like to thank Program Study Geology and Department of Geography Universitas Indonesia for supporting this research. This research was funded by Universitas Indonesia International Indexed Publication (PIT-9) grant with contract number NKB-0035/UN2.R3.1/HKP.05.00/2019.

References
[1] Jati, R. dan Amri, M.R., 2016. Indonesian Disaster Risk. Jakarta: Badan Nasional Penanggulangan Bencana. [in Bahasa]
[2] Liu, J., Doan, C.D., Liong, S.Y., Sanders, R., Dao, A.T. dan Fewtrell, T., 2015. Regional frequency analysis of extreme rainfall events in Jakarta. Natural Hazards, 75(2), pp.1075-1104.
[3] Putuhena, W.M. and Ginting, S., 2013. Development on Jakarta Flood Modelling System. Jurnal Teknik Hidraulik, 4:1, pp. 63-78. ResearchGate.net.
[4] Tehrany, M.S., Pradhan, B., Mansor, S. dan Ahmad, N., 2015. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. Catena, 125, pp.91-101.
[5] Kazakis, N., Kougias, I. dan Patsialis, T., 2015. Assessment of flood hazard areas at a regional scale using an index-based approach and Analytical Hierarchy Process: Application in Rhodope–Evros region, Greece. Science of the Total Environment, 538, pp.555-563.
[6] Shafizadeh-Moghadam, H., Valavi, R., Shahabi, H., Chapi, K. dan Shirzadi, A., 2018. Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. Journal of environmental management, 217, pp.1-11.
[7] Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F. dan Mosavi, A., 2019. An Ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Science of the Total Environment, 651, pp.2087-2096.
[8] Darabi, H., Choubin, B., Rahmati, O., Haghighi, A.T., Pradhan, B. dan Klove, B., 2019. Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques. Journal of Hydrology, 569, pp.142-154.
[9] Turner, Peter, 1997. Java (1st edition). Melbourne: Lonely Planet Publications. pp. 37
[10] Turkandi, T., 1992. Geological Map of Jakarta and the Thousand Islands, Java 1: 100,000 scale. Bandung: Dit. P3G, Dit. Geologi dan Sumberdaya Mineral. [in Bahasa]
[11] Stockwell, D., 1999. The GARP modelling system: problems and solutions to automated spatial prediction. International Journal of Geographical Information Science, 13:2, pp.143-158.
[12] Loh, W.Y. and Shih, Y.S., 1997. Split Selection Methods for Classification Trees. Statistica Sinica, 7, pp.815-840.