

# Application of gis on determination of flood prone areas and critical arterial road network by using chaid method in bandung area

Darwin<sup>1</sup>, Benecditus Kombaitan<sup>2</sup>, Gatot Yudoko<sup>3</sup>, HeruPurboyo<sup>4</sup>

<sup>1</sup>PhD Student of Transportation, SAPPK, Institut Teknologi Bandung, Bandung, Indonesia

<sup>2</sup>Urban and Regional Planning, SAPPK, Institut Teknologi Bandung, Bandung, Indonesia

<sup>3</sup>School of Business and Management, Institut Teknologi Bandung, Bandung, Indonesia

<sup>4</sup>Transportation, SAPPK, Institut Teknologi Bandung, Bandung, Indonesia

**Abstract.** Floods in Bandung area often occur when the rainfall is high then the water volume exceed the capacity of Citarum watershed. Floods cause economic and social losses. The purpose of this research is to get the GIS application model in the estimation of puddle area and road network in Bandung Metropolitan Area has disturbed. Geospatial map preparation methodology used statistical data from 11041 flood points, which divided into two groups, 7729 flood points to estimate the decision tree model and 3312 flood points to validate the model. The process of making flood vulnerability maps is approached by Chi-square Automatic Interaction Detection (CHAID) method, and validation using Receiver Operating Characteristic (ROC) method. Validation results in the area under the curve with a value of 93.1% for success rate and 92.7% for the prediction level. Chaid result is class 0 - 0,047 covering 76,68% area; Grades of 0.047-0.307 include 5.37%; Grades 0.307 - 0.599 (Low) covering 5.36%; Grades 0.599 to 0.4444 include 5.31% and grade 0.844-1 (high) covering 7.27% of the research area. Flood-prone road network is Link from Rancaekek (Area of PT Kahatex), link from Solokan Jeruk (Cicalengka-Majalaya), Link Baleendah, and link Dayeuhkolot (M.Toha - Andir)

## 1. Introduction

Indonesia has 17,508 islands located between two continents (Asia and Australia) and between two oceans (Indian Ocean and Pacific). As the world's largest archipelago country, Indonesia, is at the meeting of three major plates of the world that is the Indo Australia, Eurasian and Pacific Plate, and 9 small plates that have the potential to cause earthquakes if the plates collide so that the Indonesian territory becomes prone to various disasters, both earthquakes, Forest fires, tsunamis, floods, landslides, storms, and disasters caused by technological failures. In addition, Indonesia also has 129 active volcanoes, 80 of which are dangerous as shown in figure 1.

The National Disaster Management Agency has recorded disaster statistics as shown in Figure 2 on the following page. Based on the statistics of disaster occurrence in Figure 2 it shows that floods are frequent, whereas earthquakes and tsunamis cause the most deaths. Flood incidents have the highest intensity, causing traffic disruption, both congestion and broken roads, and bridges damage, housing, and some shops and industrial activities are disrupted hence causing huge economic losses. This condition as a result of

global climate change hence the incidence of this flood will not decrease in the near future, [11].

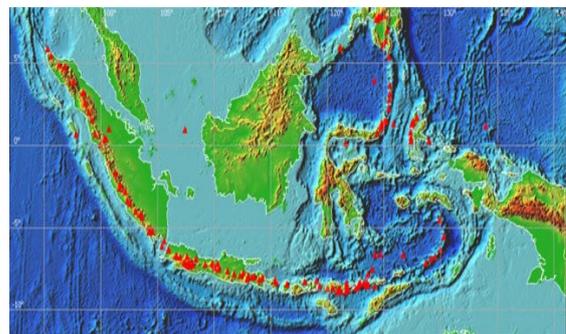


Figure -1. Active volcanoes Distribution in Indonesia

Bandung metropolitan area covers 71 districts, especially in the area of South Bandung which experienced flood disaster every rainy season, as shown in figure- 3 in the following page.

The major floods of 2010 in southern Bandung covered an area of 6,420 ha and caused economic losses of up to Rp 615 billion [5] and the floods in 2016 more bigger than 2010.

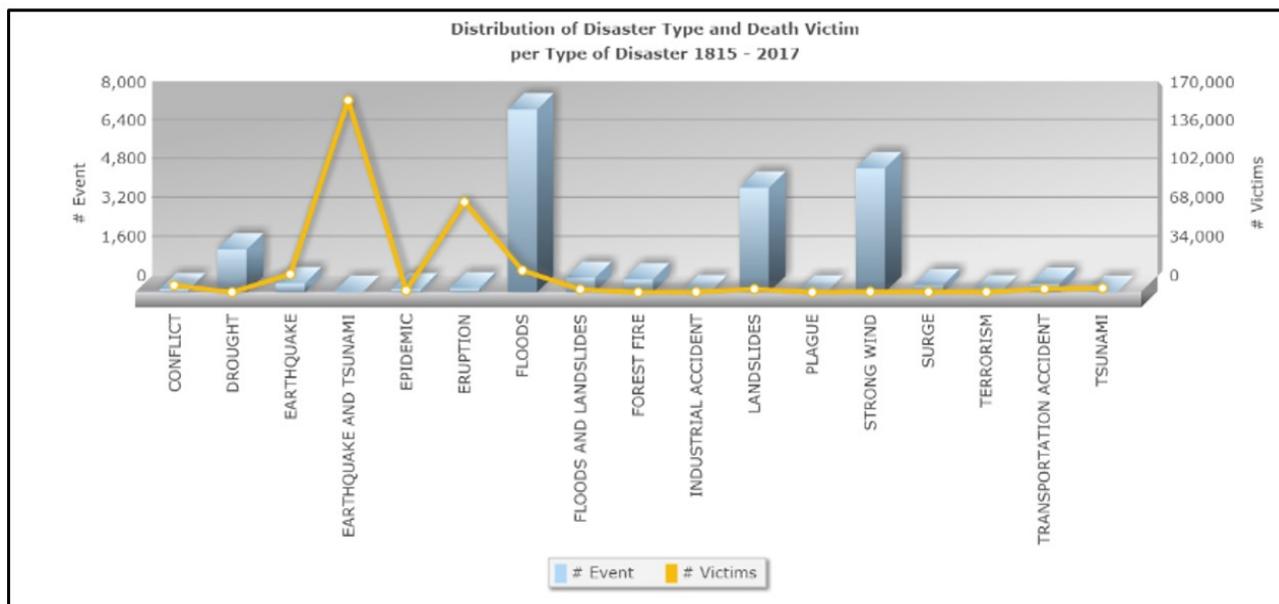


Figure - 2 The statistics of Indonesian disaster occurrence



Figure – 3 Floods on arterial roads in Bandung Area

Flood modeling (including hydrological process modeling) is carried out with various techniques, including rainfall modeling techniques - rain runoff, data-based techniques, and a combination of both [12]. Physical modeling techniques are more difficult because of the dynamics of change in water catchment areas. There is a model that requires an understanding of the physical processes occurring in the watershed so that the data-driven approach to the flood model is increasingly explored [16].

Theory of machine learning methods to be the primary choice for disaster modeling with data-driven approach. Decision tree method derived from the theory of machine learning is a highly efficient tool for classification and estimation. Unlike other statistical methods, decision tree methods do not take statistical assumptions, capable of handling data that is represented by the various measurement scale, and it only takes a short computation [15].

The utilization of GIS for flood mapping is conducted in Jambi [7], for green open space in Depok

area [7], but they do not discuss the flood-affected road network. In the Bandung area there has been no research on floods and road networks that are affected so as to disrupt traffic flow. The analysis of traffic disturbances caused by flooding is very important in transportation planning, especially to conduct traffic engineering so that greater losses can be avoided.

The purpose of this research is to get the GIS application model to estimate flood area and road network disturbed by flood. Flood analysis model using advanced decision-based rule tree by utilizing Geographic Information System (GIS) software to process thematic data and then using module that provided by SPSS software to generate decision tree. Then decision trees are used to generate flood prone areas and re-use GIS software to illustrate flood-prone maps. The benefit of this research is to obtain a map of flood-prone areas in Bandung so that the government can conduct good disaster management, such as planning of evacuation system, flood logistic needs, and traffic engineering management.

## 2. Literature Review

### 2.1. Research Area

The area of Bandung metropolitan includes 71 districts, which entered into 5 cities, namely Sumedang District, Bandung City, Cimahi City, Bandung Regency, and West Bandung. The shape of Bandung's highlands consists of a plain of about 40 km from west to east and 30 km from north to south, with a population of 7 million in 2005 and is projected to increase to 12.8 million by 2025, while the city of Bandung population is 2,470,802 People in 2014 [1].

The geological conditions of the Citarum watershed are divided into several morphologies such as those in the Atlas of Water Resources of the Department of Water Resources Management (PSDA).

Upstream watersheds in the Citarum Basin are formed from volcanic morphology with relief of soil and gravel, with a height of 750 - 2300 m above sea level, intersected 5-15% feet, in the middle of 15-30. %, and at the peak of 30-90%. The pattern of parallel river flow and river flow is the main sink area of shallow air and in the air where the slope of the slope is released. Stone constituents are the sediments of young and old volcanoes, which consist of tuff, breccia, and lava.

The river flows in Bandung quite a lot, such as the river Cikapundung, Cidurian, Cinambo, Cisangkuy, and others that empties into Citarum. In extreme weather conditions floods often occur because the capacity of drainage and Citarum rivers is lower than the volume of rainwater.

### 2.2. Geographic Information Systems

Geographic Information System is an application designed to perform various operations on geographic information. Geographic information is defined as information about the location on or near the earth's surface, and can be organized in various ways [4]. GIS presents the information in a graphical form by using the map as an interface. GIS is composed of the concept of multiple layers (layer) and relationships. Each layer in GIS represent data and specific information according to the geographical location and the relationships that are defined [6]

Geographical phenomena is generally conceptualized using continuous fields and discrete objects. But these two concepts problematic when digitized as a result of the information contained therein was not clearly quantified. So if you want to describe a geographic phenomena that can be processed by a computer, the format should be changed using the two methods, the vector and raster. Basically vector and raster can be used to codify the continuous field and discrete objects, but there is a strong correlation between raster and continuous field and vector with discrete objects [13]

GIS database is usually conceptualized as a collection of layers, each layer bound to the coordinates on the earth's surface. A layer can contain representations

of fields, such as land cover maps, or may contain multiple discrete objects, such as buildings. Layer-based raster usually describe only the geographical variation of one property only, such as the name of the district or land cover types, so the raster layers usually conceptualized with a continuous field. Layer-based vectors on the other hand can represent a collection of discrete objects with a number of diverse attributes, or a variation of a property which is conceptualized by a continuous field [4]. Database in GIS application known as the "geodatabase" because the data in the database of this type are usually equipped with geographic coordinates.

### 2.3 Decision Tree and CHAID Algorithm

The decision tree is a data mining technique that breaks down the heterogeneous data set into the group of more homogeneous data by using directed knowledge discovery. Directed knowledge discovery explains the relationship between the target variable with other variables to find certain patterns, which are then used to predict future events by using the chain of decision rules. Chain of decision rules generated using the patterns found in the analyzed data. In this way, a decision tree model provides an accurate and clear where a decision tree model can explain the reason of certain decisions using decision rules. The decision tree can be used in classification problems and also in the estimation problem in which the output is a continuous value.

CHAID stands for "Chi-Squared Automatic Interaction Detection". This method follows the three stages in the decision tree generation process, namely: merging, splitting, and stopping. As the name implies, CHAID using a p-value of the chi-square as merging and splitting criteria. Splitting the CHAID method was not consider limited the binary split, but also consider multi-way split. CHAID algorithm is able to receive input nominal and ordinal and able to accommodate different scale measurements of each variable

## 3. Research Methods

### 3.1. Model GIS

relates to the standardization of spatial maps. The coordinate system used is WGS 1984 UTM Zone 48S with Transverse Mercator projection, and with the geographic coordinate system GCS WGS 1984 datum D WGS 1984. The Linear units used in this coordinate system is meter.

Geographical Information Systems (GIS). GIS is a generic term denoting the use of computers to create and depict digital representations of the Earth's surface. From humble beginnings in the 1960s, GIS has developed very rapidly into a major area of application and research, and into an important global business. [14]. GIS models used in this study

### 3.2. Decision Tree

There are various models of decision tree that can be used to process the data in this study. Models of the decision tree include: Classification and Regression Tree (CART), Quick, Unbiased, Efficient Statistical Tree (QUEST), Chi-square Automatic Interaction Detector (CHAID) and C4.5 algorithms. The model used in this study are CHAID because CHAID analysis able to accommodate categorical data and continuous data at a time. This factor is suitable for use in the data in this study, because there is a categorical data (soil type, land cover, and rainfall) as well as continuous data (altitude, slope, curvature, and the distance from the river). Besides, CHAID analysis is not limited to do a binary split only, but be able to do the splits with the result more than two nodes. CHAID analysis thus produces a decision tree nodes that are more specific.

### 3.3. Research Steps

GIS data processing carried out in accordance with the model analysis described in next section. Processing begins with a projection of spatial maps into the selected coordinate system. The first spatial map projected is DEM because these maps will be used as a base map to facilitate the processing of other spatial maps. Projections into the coordinate system and then proceed to map soil types, land cover maps, maps Citarum river and tributaries Citarum, rainfall map, a map of past floods, and a map of the study area. Then do cutting the area of spatial maps with a map of the study area Metropolitan Bandung Area. This is done to focus the map region be analyzed only in the area of research only. The next stage in the form of DEM data processing into slope map and curvature maps.

Then the whole spatial maps that are in vector format (maps of soil type, land cover maps, rainfall maps, map distance from the river, and maps of past flood) is converted into a raster format. The process of conversion from vector to raster or rasterization is done by the method of maximum combined area. The entire map in vector format is converted into a raster format with a resolution of 92.72 m, the same as the map resolution DEM.

Map of flooded areas that have been in rasterization is divided into two, with the division of 70% of cells and 30% of the cells. Cell division is done at random and will be used in conjunction with other data at the stage of processing the decision tree. Other data to be used for processing the decision tree is a data cell that is not subject to flooding of a number of cells flooded area. The cells was not affected by the floods also were randomly divided into sections of 70% and 30%.

The generation of a decision tree using a data from extraction result at the processing stage GIS. 70% of cell data is used to perform decision tree training and 30% cell data is used to perform testing a decision tree as a result from generation. Before performing the generation of a decision tree, there are several criteria that must be set. These criteria is the dependent variable, independent variables, limiting criteria, level of significance, and the size of the independent variable interval. This phase will

generate decision trees along with the rules of each terminal node in the decision tree. These Rules along with the data overlaying the entire spatial map was used in the processing stage flood-prone maps.

Rules node terminal as a result from generation decision tree along with each overlaying data cell from GIS process become an input to generate flood-prone maps for the Metropolitan area of Bandung Area. Rules node terminal are used to generate the probability of flood-prone in every overlaying maps cell. The probability of flooding already known is then used to visualize them in the form of a map using GIS software. Flood probability map is then classified into several classes to determine the level of danger of flood-prone throughout the Metropolitan area of Bandung Area.

## 4. Data Collection

The data used in this research is secondary data collected from various sources. Here is described a method of collecting individual data points.

Flood maps is obtained by processing remote sensing imagery or aerial photographs that were recorded during the flood. Using algorithms, to identify the surfaces that are flooded from the recorded image. The results of identification using algorithms then check back to verify the entire surface of the flood. The map of past Flood shown in figure 4.

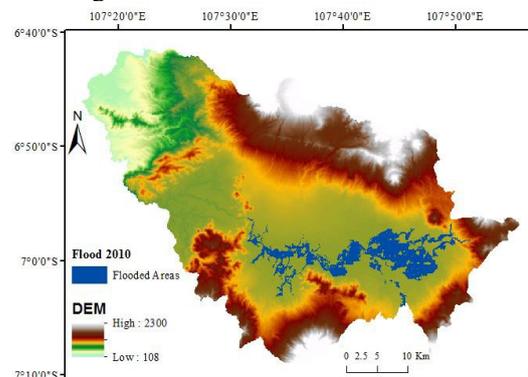


Figure 4. Flood in the past

Map land cover was obtained in a manner similar to the flood maps, which is by processing the remote sensing imagery or aerial photography. The algorithm used to identify the type of land cover more advanced because the object to be identified more as compared to only one object for identification flooding. Just as the flood maps, to map land cover, need to be verified by the man against the identification is done by algorithms and the result shown in figure 5.

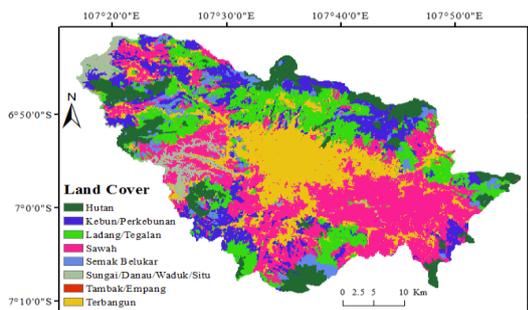


Figure 5. Covering Area

Rainfall map was obtained by collecting rainfall data from various measuring stations rainfall in the region studied. Location of rainfall stations accurately mapped using the coordinates and is associated with rainfall data. Then the rainfall data and the location of measuring stations are interpolated using Kriging method. Interpolated into the rainfall map is used as shown in figure 6.

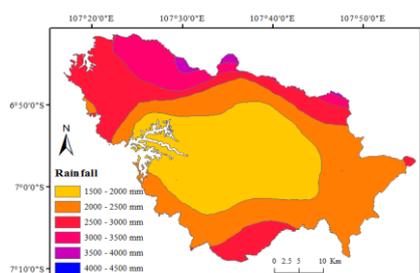


Figure 6. Rainfall map

Map of soil types is obtained by performing soil sampling at locations spread out evenly. Soil sample data along with the coordinates of the sampling is stored for processing. The data is then interpolated using Kriging method, resulting in a soil type map for the area sampling as shown in figure 7.

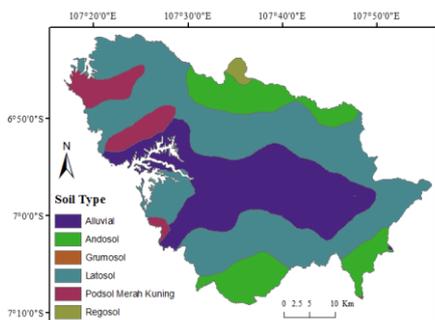


Figure 7 Soil Type

DEM data source is from the National Aeronautics and Space Administration (NASA) generated by an international study known as the Shuttle Radar Topography Mission (SRTM). SRTM data in the form of topographic world map with a resolution of 90 m. Data results are distributed by NASA SRTM have holes

where the pixels have the value "no-data". So the data must be cleaned with one of them re-interpolation. Data cleansing is done by Jarvis [10] and the results of such processing is open to the public. Data is distributed by the Consortium for Spatial Information (CGIAR-CSI). DEM map for the area of West Java and the surrounding shown in DEM map with latitude 10S - 5S and longitude 105E - 110E.as shown in figure 8

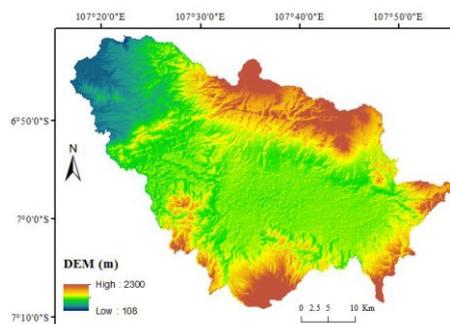


Figure 8. DEM

Slope and curvature maps of the soil surface is obtained by performing data processing DEM. Map of the slope is the first derivative of DEM map, and map the curvature is a derivative of the slope or the second derivative map from DEM map.

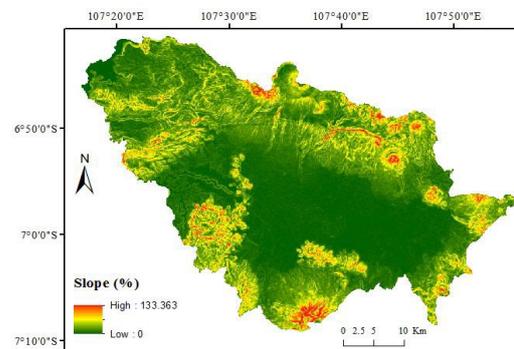


Figure 9. Slope of soil surface

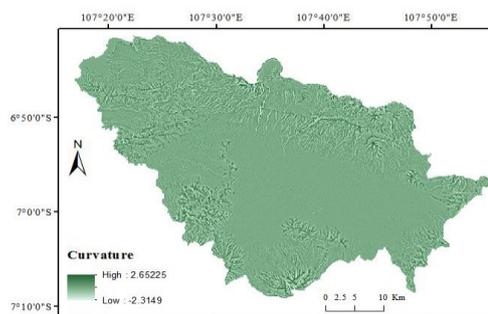


Figure 10. Curvature of soil surface

Map of the Citarum river and tributaries Citarum obtained by making use of remote sensing imagery or aerial photography. Citarum river imagery is used to determine the boundaries of the river. The boundaries of the river described using appropriate lines and then saved as a polyline features in the GIS software for further processing.

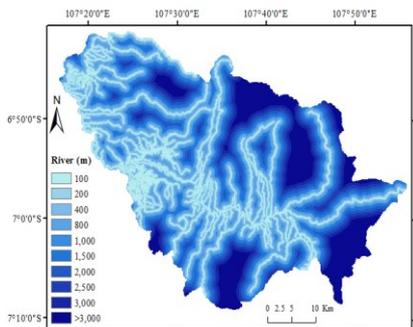


Figure 11. Distance from River

Map of the road network in the Metropolitan Bandung Area obtained from non-profit institutions Open Street Map (OSM). The institute aims to provide a complete map of the whole world with open access. Figure polyline path obtained by processing remote sensing imagery or aerial photography.

All data obtained are fed into statistical input data to be processed with Chaid to obtain the probability that an area will experience floods based on past experience. The data in the input format of the statistics processing is shown in the figure below.

All XYPoints.sav [DataSet1] - IBM SPSS Statistics Data Editor

	OBJECTID	DEM	Slope	Curvature	LandCover	SoilType	Rainfall	Flood	River	POINT_X	POINT_Y
1	1	110	1.348190	.011633	8	4	1	0	100	762194.799970000	9259932.80955000
2	2	327	6.092250	-.174490	8	4	2	0	100	766181.639083000	9259932.80955000
3	3	337	5.382620	.046531	5	4	2	0	100	766274.356272000	9259932.80955000
4	4	338	4.181550	-.058163	5	4	2	0	100	766367.073460000	9259932.80955000
5	5	335	6.346510	-.162857	5	4	2	0	100	766459.790649000	9259932.80955000
228895	228895	1881	20.27310	.348980	1	1	1	0	10000	787877.461230000	9208938.35578000
228896	228896	1891	22.73040	.255919	1	1	1	0	10000	787970.178419000	9208938.35578000
228897	228897	1914	27.14110	.442042	1	1	1	0	10000	788062.895607000	9208938.35578000
228898	228898	1937	27.22730	.186123	1	1	1	0	10000	788155.612796000	9208938.35578000
228899	228899	1965	28.71410	-.034898	1	1	1	0	10000	788248.329985000	9208938.35578000
228900	228900	2002	26.92320	-.174490	1	1	1	0	10000	788341.047173000	9208938.35578000
228901	228901	2029	14.50160	.314082	1	1	1	0	10000	788433.764362000	9208938.35578000

Figure 12 Input Data Statistik for Chaid analysis

### 5. The results of the study

There are criteria for the generation of a decision tree was defined in this study. The dependent variable was the incidence of floods (**Flood**) or no flood (**No Flood**). The independent variable is DEM, rainfall, soil type, distance from the river, slope and curvature. Limiting criteria used are minimum of 100 cases for parent nodes and a minimum of 50 cases for child nodes. The significance level (alpha) is 0.05 or 95% confidence

level. From the results of the process output module Decision Trees on SPSS software with the method of calculation CHAID produced decision tree (Figure 13) with overall 98 nodes and 57 terminal nodes. All independent variables are acceptable for use in generating the decision tree and decision trees have maximum depth of as much as 6 levels.

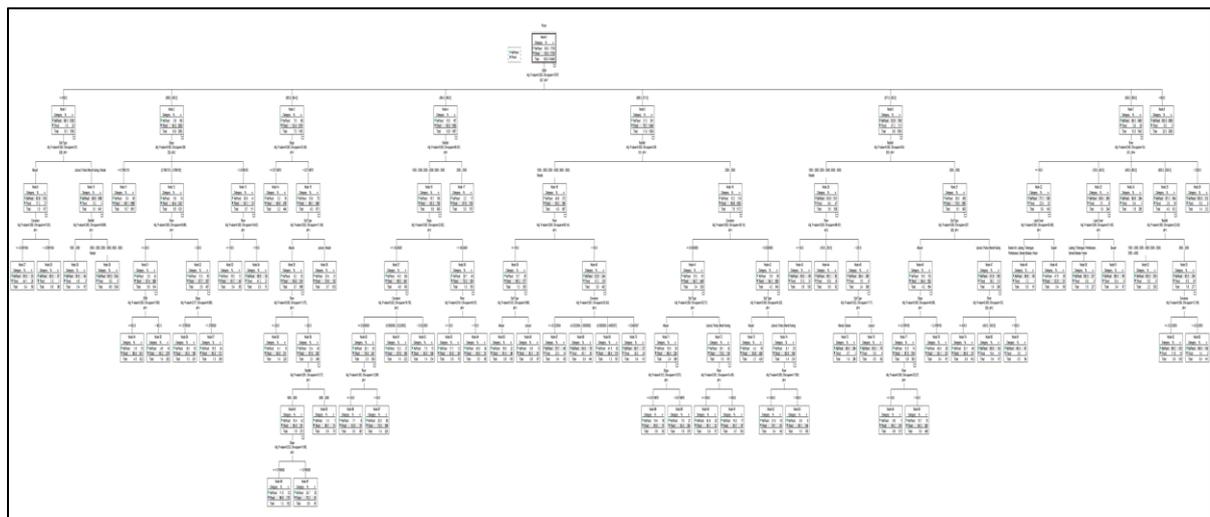


Figure 13 The Tree of flooding probability

The decision tree is generated to produce rules that are then used to analyze the new data set. There are 57 pieces of Rules decision tree, this rules are used to create SPSS syntax and then will be used in determining the probability of flooding. The model validation is done by using the method of Receiver Operating Characteristic curve (ROC). Validation is done by using the success rate curve and the curve prediction rate. Success rate curves are generated using 70% the training data set (7729 points floods). Success rate curve produces the value Area Under the Curve (AUC) which is equal to 0.931, or representing the success rate accuracy of 93.1%. While the prediction rate curves are generated using 30% data sets of the flood point (3312 points) that have been saved to do the validation. Prediction rate produces curve value (AUC) which is equal to 0.927 or representing the prediction rate accuracy of 92.7%. The final result of the generation of the flood vulnerability maps based on the rules derived from a decision tree in the form of a map of the probability of flooding with a range of values of 0 to 1. Where the closer a number to 1, the greater the probability of flooding and otherwise getting close a number to 0, the smaller the probability of flooding.

In the making of flood vulnerability maps in this study, there are four types of classification methods considered for use. Fourth methods such as quantiles,

Map of flood vulnerability of this research is beneficial for a variety of things. One was to identify the main street in the Metropolitan Bandung Area who have a high risk of flooding affected. This analysis is done by using the main road map in overlaying position to the maps of vulnerability flooding.

The duration of traffic disturbance depends on the time of flood receding. From the results of interviews with flood-affected communities, the following regression equation is obtained:

$$Y = 8.830 + 0.575X1 + 28.524X2 \quad (1)$$

Where:

natural breaks, equal intervals, and standard deviation. Then each method tested and selected the most appropriate method to the scope of this study and with the kind of information wish to display.

Natural breaks method suitable for use if there is a significant leap between the values of which are classified [2]. In this study, there were significant leap between the probability value was not flood with a probability value flooding, so use natural breaks classification method shown in Table 1

Table 1. Classification of the probability of flooding area

Probability	Class	%
0 – 0.047	Extremely low	76.68
0.047 – 0.307	Very low	5.37
0.307 – 0.599	Low	5.36
0.599 – 0.844	Moderate	5.31
0.844 - 1	High	7.28

Y :Predicted flood time recedes (hours)

X1: Duration of rainfall with intensity > 60 mm in Bandung area (hour)

X2: Depth of flood (m)

The coefficient of determination  $r^2 = 0.734$ , which shows that the equation can only explain as much as 73.4% that the length of time recedes is influenced by the depth of flood and the period of heavy rainfall in Bandung area..Traffic disturbances occur during flooded roads. Losses resulting from traffic congestion and lost productive time and decreased vehicle utility. In addition to the loss of delay in delivery of goods, especially for logistics vehicles. In this study not yet calculated all social losses, but the duration of traffic

disturbance is indicated by the length of time receded in the arterial road until the road can function again.

## 6. Discussion

The decision tree results from the data processing has 6 levels of depth and is very wide relative to the side, due to the many separations in some independent variables. The land altitude variable becomes the most significant variable in flood event since it has the smallest p value adjusted to the root node among other variables. Thus the land altitude variable becomes the first branch of the root node. As the first branch of the root node, variable land altitude automatically become predictors for all terminal nodes, thus always affecting flood prediction or not.

There are several things that should be discussed in the model produced in this study. Model decision trees and model prediction capabilities are strongly influenced by training data used at the start of the study. Moreover, training data is an example of historical data that is static. So if using different training data, decision trees and predicted results are likely to change. So the decision tree model is less powerful. This is one of the major disadvantages of Chaid's decision tree method. Due to this deficiency, this model is only suitable for making predictions according to current circumstances. Predictions of future floods with certain rainfall could be done, but it is feared the results are not accurate.

The next thing is related to the decision making process. The decision tree making algorithm that applies to SPSS software states that flood or non-flood categorization depends on the percentage of most terminal nodes. Cutting categorization is greater than 50%. So if one category value is greater than 50% on the terminal node, then the category will be the flood status in each terminal node.

It should be remembered that the model developed in this research is a simplified model of physical flood model that tends to be very complex. So the flood vulnerability map from this study can be considered as a simplification of the actual flood vulnerability. But still an accurate prediction based on the data source used to make predictions. The advantage gained is that it simplifies the modeling time much faster thanks to the data-driven approach and the ease of data collection, which only uses the existing secondary data.

When pay attention to Figure 14 in the form of flood prediction and 4 in the form of the past flood area, looks rough correlation between the incidence of past flooding with the flood prediction. So in the visible, it looks a relatively strong correlation between predictions of flooding to the real conditions are illustrated on the maps spatially. Seen a few predictions of floods that shaped the line snaking out of the main collection. The lines illustrate the vulnerability of flood higher on the banks of rivers region.

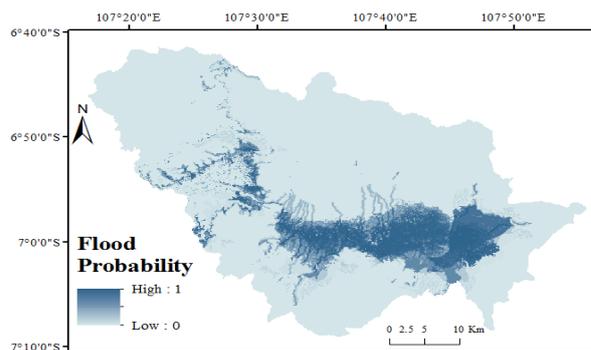


Figure 14. Map of the probability of flooding

Figure 15 shows that the flood vulnerability map dominated by classification extremely low (very low). When pay attention to Table 1 shows that the classification of the extremely low probability ranges from 0 to 0.047 covered 76.68% of the whole area which is reflected in Figure 15. It is far different from the four other classification with details very low 5.37%, low 5.36% , moderate 5.31% and 7.28% high. This indicates that most of the Metropolitan area of Bandung Area has a very low vulnerability of flood. Flood events are concentrated in certain areas only.

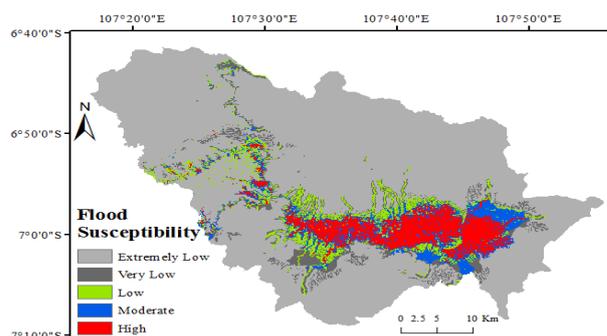


Figure 15. Map of flood vulnerability

Figure 16 shows an arterial road network that can experience flooding during the rainy season, so the road is susceptible to flooding. The analysis does not include the possibility of the flood water level, so it can not be determined which path is completely disconnected or just stagnant. Figure 17 shows a map of the Bandung area, with some arterial road names, and combined with Figure 16 it can identify potential critical road names when floods arrive.

The model developed in this research still has some shortcomings, mainly the set of data used to have differences in the year of collection. The development of future research should start by updating the sets of data with new data sets. Besides the accuracy of pixels should be increased in order to improve the accuracy of predictions resulting flood vulnerability maps are much more accurate. Suggested accuracy used was 15 m x 15 m or 10 m x 10 m. Accuracy like this can allow the use of data Stream Power Index (SPI) and Topographic wetness index (TWI) which requires precision pixels high.

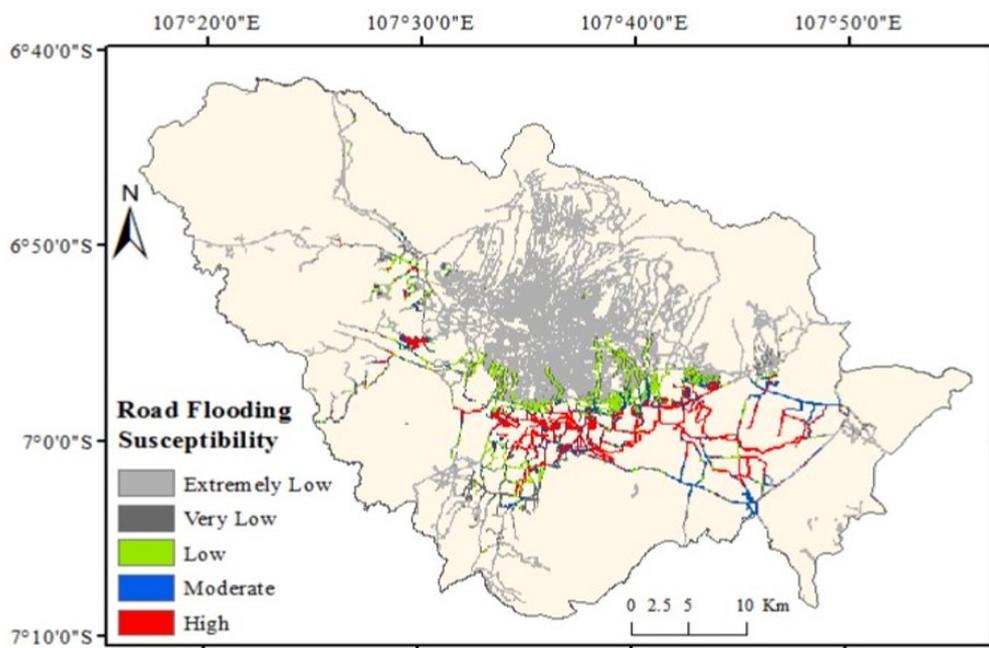


Figure 16. Map of road network flood vulnerability

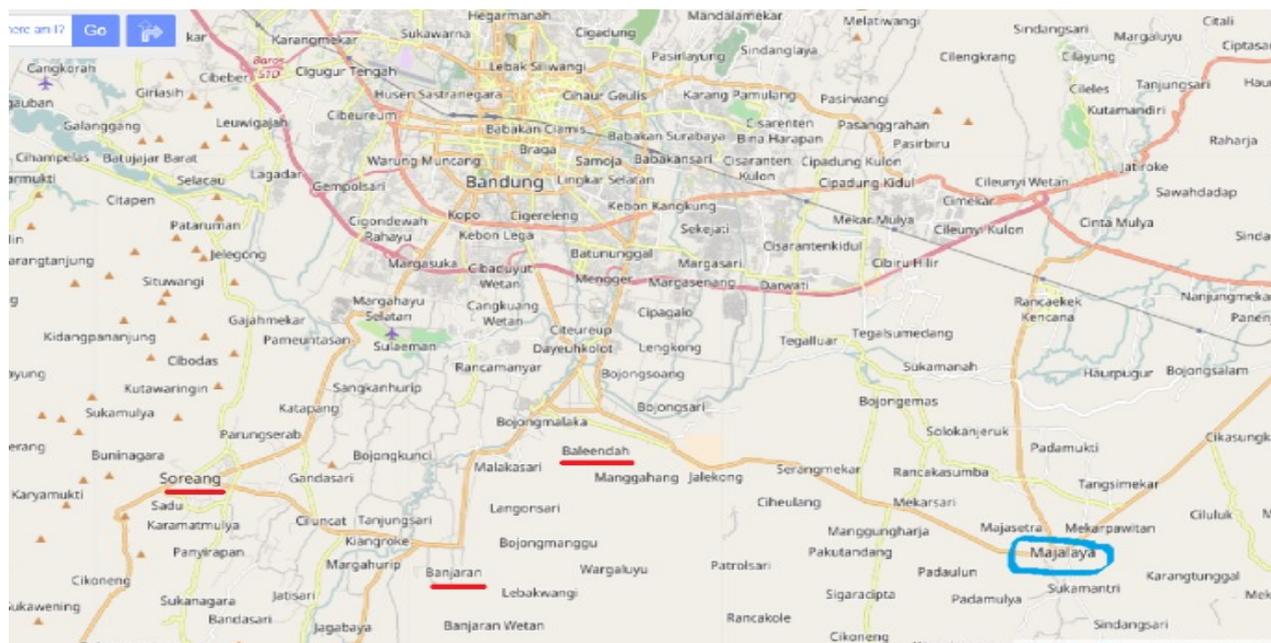


Figure 17. Map of the road network with identity

## 7. Conclusion

Preparation of the Bandung area highway map is done by using the application of geographic information system. Flood event data in that area was processed by using CHAID classification analysis obtained by flood risk map, and then overlaid with Road Network to get flood road network map.

Digital Elevation Model (DEM) data, land cover, rainfall, soil type, distance from Citarum river, slope and curvature, Road identification was analyzed by Tree Diagram model and CHAID resulted in decision tree,

which has 97 nodes, and 57 nodes Terminal node. A possible map of flood is generated using rules based on the terminal node. In predicting the flood, the DEM variable shows the most significant. The validation process of decision tree prediction model shows identification of 93.1% flood point with prediction rate of 92.7%. Classification of flood-prone areas obtained by Chaid result are class 0 - 0,047 covering 76,68% area; Grades of 0.047-0.307 include 5.37%; Grades 0.307 - 0.599 (Low) covering 5.36%; Grades 0.599 to 0.4444 include 5.31% and grade 0.844-1 (high) covering 7.27% of the research area. Based on the result of

overlay map processing, the critical arterial road network is the road segment of Rancaekek - ParakanMuncang (Area of PT Kahatex), Solokan Jeruk(link ofCicalengka - Majalaya). Baleendah (Link of Bojongsoang - Baleendah), and link ofMochamadToha - DayeuhKolot, which will experience flooding if it rains more than 3 hours in all areas of Bandung. The flood vulnerability maps in this study make it easier for stakeholders to identify areas that need more attention in reducing flooding.

## 8. Suggestion

Further research is recommended to use the higher pixel precision such as 15 x 15 m, or 10 x 10 m in order to increase the accuracy of the data processing. In addition, higher data accuracy allows the use of SPI and TWI variables in research.

Additionally further research may use a decision tree method such as CART method, QUEST, C4.5 or other machine learning methods like ANN, Logistic Regression, or Naïve Bayesian. Then the results were compared with the results of this study to analyze the accuracy of each method in predicting flood vulnerability in the Metropolitan Bandung Area. The research results can be used to analyze other data that associated with vulnerability to flooding. Such as calculating the potential loss due to floods, the number of vulnerable populations affected by the floods, the area of agricultural potential damaged by flooding, and a variety of other uses.

Floods and landslides occur every year, especially in the rainy season, therefore the Regional Disaster Management Agency needs to redesign the disaster database system, by incorporating 3-dimensional coordinates on the geographic information system map so as to predict the potential of flood areas, network of critical road, time recede of flood and time of traffic disruption. Therefore it is necessary to coordinate disaster mitigation with Dinas Bina Marga and DinasPerhubungan and Police to develop contingency plan in case of flood disaster so that social losses can be minimized by educating the public to work together to prevent flooding earlier and to conduct traffic engineering management.

We would like to thank Mr. Oki (Pusair) and Mr. Dedi (BWS Citarum) for assisting in providing data on the characteristics of Citarum and ZhilalYusa that have assisted the preparation of GIS maps and the flood affected communities who have been willing to be interviewed to provide information on flood behavior in the area of Bandung

## 9. Reference

1. Abidin, H. Z., Gumilar, I., Andreas, H., Murdohardono, D., & Fukuda, Y. (2013). On causes and impacts of land subsidence in Bandung Basin, Indonesia. *Environmental earth sciences*, 68(6), 1545-1553.
2. Ayalew, L., & Yamagishi, H. (2005). The application of GIS-based logistic regression for

- landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology*, 65(1), 15-31.
3. Dam, M. A. C., Suparan, P., Nossin, J. J., & Voskuil, R. P. G. A. (1996). A chronology for geomorphological developments in the greater Bandung area, West-Java, Indonesia. *Journal of Southeast Asian Earth Sciences*, 14(1), 101-115.
4. Goodchild, M. F. (2009). Geographic information system. In *Encyclopedia of Database Systems* (pp. 1231-1236). Springer US.
5. Gumilar, I., Abidin, H. Z., Hutasoit, L. M., Hakim, D. M., Andreas, H., Sidiq, T. P., & Gamal, M. (2012). Pemetaan karakteristik penurunan muka tanah berdasarkan metode geodetic sertadampaknyaterhadapperluasanbanjir di cekungan. *MajalahIlmiah Globe*, 14(1).
6. Hanifah, R., Isnanto, R. R., & Christyono, Y. (2010). Simulasisisteminformasigeografis (SIG) pemantauanposisi kendaraan via SMS gateway. *TRANSMISI*, 12(2), 45-49.
7. HendiHamdani, SulwanPermana, AdiSusetyaningsih (2014), Analisis daerah rawan banjir
8. Menggunakan aplikasi sistem informasi geografis (studikasuspulaubangka), *JurnalKonstruksi*, ISSN: 2302-7312 Vol. 12 No. 1 2014.
9. Irina Mildawani, Diana Susilowati, Lia Rosmala Schiffer (2016), Aplikasi sistem informasi geografis (sig) Dalam analisis pemanfaatan dan pengelolaan Ruang terbuka hijau kota (rthk) Studikasu: kotadepok, Lembaga Pengembangan Fakultas Teknik Sipil dan Perencanaan (LePTESP), Universitas Gunadarma.
10. Jarvis A., H.I. Reuter, A. Nelson, E. Guevara, (2008). Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT)
11. Khan, S. I., Hong, Y., Wang, J., Yilmaz, K. K., Gourley, J. J., Adler, R. F., ... & Irwin, D. (2011). Satellite remote sensing and hydrologic modeling for flood inundation mapping in Lake Victoria basin: Implications for hydrologic prediction in ungauged basins. *IEEE Transactions on Geoscience and Remote Sensing*, 49(1), 85-95.
12. Kia, M. B., Pirasteh, S., Pradhan, B., Mahmud, A. R., Sulaiman, W. N. A., & Moradi, A. (2012). An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. *Environmental Earth Sciences*, 67(1), 251-264.
13. Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). *Geographic information systems and science*. John Wiley & Sons Ltd.
14. Paul A Longley, Michael F goodchild, David J Maguire and David w rhind, 1999, *Geographical information system, Principles and Techniaal issues*, 2<sup>nd</sup> edition, Jhon Wiley.
15. Saito, H., Nakayama, D., & Matsuyama, H. (2009). Comparison of landslide susceptibility based on a decision-tree model and actual landslide

occurrence: the Akaishi Mountains, Japan.  
*Geomorphology*, 109(3), 108-121.

16. Varoonchotikul, P. (2003). *Flood forecasting using artificial neural networks*. CRC Press