Visualization of spatial decision tree for predicting hotspot occurrence in land and forest in Rokan Hilir District Riau

Aji Primajaya¹*, Imas Sukaesih Sitanggang², Lailan Syaufina³

¹ Department of Informatics Engineering, Faculty of Computer Science, University of Singaperbangsa Karawang, Karawang 41361, Indonesia
² Department of Computer Science, Faculty of Natural Science and Mathematics, Bogor Agricultural University, Bogor 16680, Indonesia
³ Department of Silviculture, Faculty of Forestry, Bogor Agricultural University, Bogor 16680, Indonesia.

E-mail: aji.primajaya@staff.unsika.ac.id

Abstract. Visualization is an important issue in datamining to easy understand patterns extracted from dataset. This research applied the Bottom-Up Approach method to develop a visualization module for a spatial decision tree in a geographic information system. Spatial data used in this work consists of nine explanatory layers and one target layers. Explanatory layers include maximum daily temperature, daily precipitation, wind of speed, distance of nearest river, distance of nearest road, land cover, peatland type, peatland depth, income source. The target layer contains hotspot and non-hotspot points that occurred in 2008. The result is the visualization module of spatial decision tree that has three main features including mapping window, interactive window, tree node and tabular visualization for predicting hotspot occurrence.

1. Introduction

Predicting hotspot occurrences is important to prevent area from fire risk. Knowledge discovery in database (KDD) is an approach used to search knowledge in large size dataset [1]. Spatial decision tree as one of methods for KDD can be applied to extract knowledge on spatial database that has high accuracy on predicting hotspot occurrence [2]. Output of spatial decision tree is rule based knowledge. On the other hand, the output is not easy to understand by users.

Visualizing output of spatial decision tree algorithm in geographical information system is essential so that users can easily use the system for predicting hotspots occurrence in a certain location. Visualization is the last step in data mining process. Visualization provides several facilities to analyze patterns and spatial relationship in spatial data distribution [3]. A Web-GIS based visualization as a combination of the web technology and the GIS technology [4] can be applied to visualize output of spatial decision trees. The system can retrieve, store, request, analyze, display of spatial data [5]. Developing the system as web engineering needs multidiscipline collaboration [6].

Several studies have been conducted on visualization of knowledge discovery in database. A spatial decision tree developed from the Iterative Dichotomizer 3 (ID3) algorithm was applied in [2] to predict hotspot occurrences in District of Rokan Hilir Riau Indonesia in which the result has higher accuracy.
than ID3. In addition, a spatial decision tree algorithm was applied on the forest fire dataset that result
the classifier with the higher accuracy than C4.5 [7]. The Spatial decision tree also has higher accuracy
than logistic regression to predict hotspot occurrences [8]. Visualization method was introduced in [9]
based on multiple trees. Tree map method was proposed to visualize spatio-temporal data in [10].
Reviewing classification of visualization method was done in [11]. The interactive map was proposed
to visualize non-spatial data in [3]. The interactive map provides several facilities that support users to
input non-spatial parameters and to interpret the output of C4.5 algorithm.

The bottom-up approach was applied in [12] to develop a web-GIS system. This approach
consisted of four steps that was used to develop a geographic information system for health exposure
assessment in Europe area [12]. This research results a system that analyze real spatial data uploaded
by experience users or non-experience users. The system was developed using open source software.
Therefore the cost of system development can be minimized. The system was tested on arsenic
concentration of Europe area for impact analysis. A web-GIS was applied in [6] to visualize the pest
life cycle simulation using the Software as Service Approaches (SaaS) method. The research
developed a system that capable to analyze impact of olive fruit fly to agricultural goods damage. It
predicts olive fruit fly life cycle. The system used PHP CRUD framework and Google Map API. A
web-GIS was adopted to visualize analysis of land suitability in the study area Yogyakarta and Central Java in [11]. The research provides spatial information for supporting agricultural businesses. A web-
GIS to visualize agricultural characteristics and the ecological system was applied in [13]. The system
has four main features including thematic map, spatial data, data analysis and decision support
facilities. Research on the development of geographic information system to help mitigate natural
hazard in Aceh has been done on [14].

This work developed a visualization module in a form of a web-GIS using the Bottom-Up
Approach method that adopted from [12] to visualize output of the spatial decision tree algorithm that
was developed in the previous work [7]. The method is composed by four main stages namely system
requirements, basic system design, software selection and system development [12]. The system
predicts hotspot occurrences in land and forest for the study area of Rokan Hilir District, Riau
Indonesia. The system was developed by open source software therefore the development cost is low.
The spatial data in the system consisted of nine explanatory layers and one target layer. Explanatory
layers consist of maximum daily temperature, daily precipitation, wind of speed, distance of nearest
river, distance of nearest road, land cover, peatland type, peatland depth, income source and the target
laye is the hotspot layer.

2. Research method

2.1. Study area and forest fire data

The research developed a visualization module of spatial decision tree for predict hotspot occurrences
in the study area Rokan Hilir district. The district was located in Riau Province where the coordinate is
between 100°16’ – 101°21’ East Longitude and 1°14’ – 2°30’ North Latitude [7].

Explanatory layers include physical layers, socio-economic layers, weather layers, peatland type
layers and peatland depth layers. Physical layers consist of river, road and land cover. Weather layers
are precipitation layer, screen temperature layer and wind speed layer. Socio-economic data is income
source. The target layer is hotspot data on the year of 2008. The data sources are provided on Table 1.
Table 1. Data and its source.

| Data                              | Source                                                                 |
|-----------------------------------|------------------------------------------------------------------------|
| Spread and coordinates of hotspot 2008. | FIRMS MODIS Fire/Hotspot, NASA/ University of Maryland                |
| Weather data 2008: maximum daily temperature, daily precipitation, wind of speed | Meteorological Climatologically and Geophysical Agency (BMKG)         |
| Physical data 2008: distance of nearest river, distance of nearest road, land cover | National Coordinating Agency for Survey and Mapping (BAKOSURTANAL)    |
| Peatland type layer, peatland depth layer | Wetland International                                                  |
| Income Source layer               | BPS -Statistic Indonesia                                               |

2.2. System development method

This study uses the Bottom-Up Approach to develop the system. The method is divided into four main steps. The steps are system requirements, basics system design, software selection and system development.

- System requirement
  The system requirement includes two tasks namely essential requirement and additional requirement.

- System design
  The system must have user friendly interface where divided into three main parts, namely spatial database, map sever, web viewer. Geodatabase is a part of the system that store spatial database. Map server connects between database and user interface or web viewer. Web viewer is a part of the system that interacts with users. The design of the system was recommended to use open source software and users do not need additional software to interact with the system.

- Software selection
  The selection of software was done based on the system design. Geodatabase needs open source software that can conduct spatial query and store the spatial data.

- Implementation and testing
  The implementation was done based on the system design and system requirements. System testing was done on this stage to evaluate the system performance to visualize of rule based knowledge.

2.3. Spatial Decision Tree Algorithm

The spatial decision tree algorithm in [2] was developed from the ID3 algorithm. It was one of several data mining algorithms to create classification models on spatial data [2]. This algorithm has higher accuracy than several methods such as non-spatial ID3 [2], logistic regression [8], C4.5 [7]. Spatial dataset for creating spatial decision tree includes explanatory layers and target layer.

There are several measures to be calculated in the spatial decision tree algorithm such as Spatial Join Relation, spatial entropy, spatial informational gain. Spatial Join Relation (SJR) is developed from Spatial Join Index (SJI). SJR is the result of features calculation on spatial relation between two layers [2]. Spatial entropy is entropy applied for spatial dataset. This algorithm needs several inputs including spatial dataset, explanatory layers, target layer and SJR [7]. This algorithm creates spatial decision tree as the output. From the tree several rules can be generated. There are 131 rules from spatial decision tree that was visualized by the system. Five examples of rules are as follows:

- Rule 1
IF Income source = Forestry AND Land Cover = Bare Land AND Wind Speed = 0 m/s – 1 m/s AND Screen Temperature = 297°K – 298°K AND Peatland Depth = very deep/very thick > 400 cm THAN Hotspot Occurrence = False.

• Rule 2
IF Income source = Forestry AND Land Cover = Bare Land AND Wind Speed = 0 m/s – 1 m/s AND Screen Temperature = 297°K – 298°K AND Peatland Depth = Moderate 100 – 200 cm THAN Hotspot Occurrence = True.

• Rule 3
IF Income source = Forestry AND Land Cover = Bare Land AND Wind Speed = 1 m/s – 2 m/s THAN Hotspot Occurrence = True.

• Rule 4
IF Income source = Forestry AND Land Cover = Shrubs AND Wind Speed = 1 m/s – 2 m/s AND Screen Temperature = 297°K – 298°K THAN Hotspot Occurrence = False.

• Rule 5
IF Income source = Forestry AND Land Cover = Shrubs AND Wind Speed = 1 m/s – 2 m/s AND Screen Temperature = 298°K – 299°K THAN Hotspot Occurrence = True.

3. Development of spatial decision tree visualization

3.1. Concept of Spatial Decision Tree Visualization
This visualization was made in the form of Web GIS that represents the knowledge as output of the spatial decision tree algorithm. It can predict potential forest and land fires based on hotspot occurrences in the district of Rokan Hilir. This ability can support decision making for preventing and detecting land and forest fire. In order to use the system, users are not required to have background of spatial analysis.

![Figure 1. Main components of Web GIS](image)

Figure 1 illustrates the main components of web-based GIS for spatial decision tree visualization. Target data and the explanatory data are stored on the spatial database. Users interact with the system using web browser and spatial data can be retrieved through map server to predict hotspot occurrences using interactive facilities in the system.

3.2. Spatial Visualization of Spatial Decision Tree

3.2.1. System requirement. Basic requirements of the system are visualize spatial data and presenting model. Additional requirements are tree node facility and tables for displaying query results of spatial data.

3.2.2. System design. The base map layer of Rokan Hilir is stored in the spatial database. The Google map that shows Rokan Hilir is displayed on the main page. The explanatory layers and the target layers can be displayed using visualization tool in the web page. The basic system design is showed on Figure 2. In general, the overall system design has two parts, namely client and server. There are
several facilities on client side that allow users to interact with the system. Users can query layers through a browser and the associated query will be forwarded to the server side. There are map server and database server in the server side. Map server will change spatial data that are stored in the database to be layers so it can be queried by the client side. The database server stores spatial data that are needed by the system. Spatial data are vector format in which the target data are point features and the base maps consist of polygon and line features. In addition, the district border layer is stored in the database. The Google base map layer is not stored on the database server but it is queried directly in the form of vector.

3.3. Software selection
This work utilizes OpenGeoSuite 3.0, OpenLayer, PHP CI, Proj4Js. OpenGeoSuite 3.0 which consists of Posgresql, PostGIS, Geoserver and GeoWebChace. There are several open source software that can be applied to develop the geodatabase such as Postgresql and MySQL. Postgresql is the open source software that can facilitate requirement of database system. Postgresql is combined with PostGIS to perform spatial data processing. The software was selected because it has larger range of spatial data processing than MySQL [12]. The software is able to perform spatial queries because it

![Architecture of the system](image_url)

**Figure 2. Architecture of the system**
has spatial analysis function for developing GIS. The software can store spatial data in tabular form and it can manage geometry data using spatial queries.

Map server can be used to connect spatial data to the internet in which it can be requested to be displayed on a Web GIS page. Geoserver is the alternate map server software that can be adopted to the system. The spatial data is connected to a map server by creating workspace and then it is uploaded to the spatial map server with certain name. The name becomes the layer name and it can be requested to display spatial data or to visualize results of spatial data analysis using a certain algorithm.

The web viewer needs software to visualize of rule based knowledge from the spatial decision tree. Several frameworks can be used to develop web viewer. PHP CI is the framework to develop an user interface. It is combined with several frameworks to make the web-GIS viewer. The other frameworks that were used are proj4js, openlayer. Proj4js is the javascript software to convert the coordinate system of layer. The layer that has certain coordinate system can be easy converted to other coordinate system by this framework. OpenLayer is the framework to construct web-GIS with several libraries. OpenLayer can be embedded to PHP CI by requesting libraries. OpenLayer provides several functions to request layers. It is collaborated with Google API.

3.4. Implementation and Testing
Implementation of the system produces three types of visualization facilities. The facilities are the map window, an interactive window, tree node and tabular visualization. Black box was used for system testing to ensure that the system is running well. Figure 3 shows the main web page of the system.

3.4.1. Map Window. The spatial data requested by users will be visualized on several parts of webpage including the map window. The map window shows the base map layer and the marker as an indicator of selected target (hotspot). The marker gives information about selected hotspot location and prediction results of the spatial decision tree. The red marker indicates true hotspot occurrence and the blue for false hotspot occurrence. Zooming scale allows user to enlarge an important area.

3.4.2. Interactive Window. The interactive window provides several features that is composed by scroll down feature, prediction button, location button, and prediction result window. Scroll down feature facilitates users to choose explanatory attributes. Prediction button is the prediction trigger. The prediction result of hotspot occurrence is showed in the prediction result window. Location button is the trigger of marker on mapping window.

3.4.3. Tree Node and Tabular Visualization. Tree node facilitates user to query all explanatory layers and the target layer. It displays query results in the tabular form and maps. Tabular visualization provides information about prediction of hotspot occurrences.
4. Results and discussion

The visualization of spatial decision tree has facilities to provide prediction of hotspot occurrences so we can determine potential high risk of forest and land fire in the Rokan Hilir. Figure 4 the visualization result of rule 117 (if distance river is low and distance road is low and income source is plantation than false) on spatial ID3 models for predicting hotspot where the visualization shows that the predicted locations are not potential hotspot.

The system development method has the similarities with [12] but there are novelty. The first is on classification method used to create the model in which the method in this study is spatial ID3 whereas in [12] 4 types of classification include equal range, equal count, natural breaks, and standard deviation. Equal method divides the data into several class with the same interval range for each class. Equal count method is a method of classification that divides the data into several classes which have the same proportion to the sample. Natural breaks method is a method where the classification is based on the grouping of making the interval on real data. Standard deviation method is a method in which the making class based on the value of the standard deviation. The second is in terms of the cases studied. In this study, statistical analysis was not performed but only testing that is done at the final stage with black box method.
Figure 4. Visualization result of a classification rule

5. Conclusion
The study produces a visualization system which is able to provide several facilities that can make visualization of hotspot occurrences prediction on land and forest in Rokan Hilir district Riau Province. The system provides output of spatial decision tree algorithm in the form of mapping window, interactive window, tree node and tabular forms. The visualization allows users to easily understand the spatial decision tree to predict hotspot occurrences based on spatial data that are stored in the database. In further research, a facility to accommodate users to input new data will be developed.

References
[1] Han J, Kamber M and Pei J 2011 Data mining : Concept and Technique 3nd edn (US: Morgan Kaufman)
[2] Sitanggang I S, Yaakob R, Mustapha N and Ainuddin A N 2013 Journal of computer science 9 244-251
[3] Andrienko G L and Andrienko N V 1999 Proceeding of the 3rd International Symposium IDA-99 vol 1642 (Amsterdam) p 747-755
[4] Zhishana Y, Run’e L, Yanjiang W and Xiaoling S 2012 International Conference on Future Energy, Environment, and Materials Energy Procedia vol 16 (Hong Kong) p 1199 – 1205
[5] Chang K. T 2008 Introduction to Geographic Information System 4th edn (New York: Mcgraw-hill)
[6] Karydis I, Gratsanis P, Semertzidis C and Avlonitis M 2013 Proceeding of the 6th International Conference on Information and Communication Technologies in Agricultural, Food and Environment (HAICTA) vol 8 (Corfu Island) p 526-529
[7] Sitanggang I S, Yaakob R, Mustapha N and Ainuddin A N 2013 TELKOMNIKA 12 p 511-518
[8] Sitanggang I S, Yaakob R, Mustapha N and Ainuddin A N 2013 Journal of applied sciences 13 p 252-261.
[9] Graham M and Kennedy J 2010 Information Visualization 9 p 235-252
[10] Slingsby A, Dykes J and Wood J 2008 Information Visualization 7 p 210-224
[11] Hazain F A, Harisno and Legowo N 2013 International Conference of Advances Science and Contemporary Engineer vol 50 (Jakarta) p 532-543
[12] Evans B and Sabel C E 2012 International Journal of Health Geographics 11 p 2
[13] Zhang H, Yi S and Wu Y 2011 22 International Conference in Energy Engineering (Bangkok) 14 p 382-386.
[14] Nasaruddin, Munadi K, Dirhamsyah M and Yuliansyah D 2011 TELKOMNIKA 9 p 89-98