TELKOMNIKA, Vol.12, No.2, June 2014, pp. 511~518
ISSN: 1693-6930, accredited A by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v12i2.2036

A Decision Tree Based on Spatial Relationships for Predicting Hotspots in Peatlands

Imas Sukaesih Sitanggang*[1], Razali Yaakob[2], Norwati Mustapha[3], Ainuddin A. N.[4]
[1] Department of Computer Science, Faculty of Natural Science and Mathematics, Bogor Agricultural University, Indonesia
[2,3] Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia
[4] Institute of Tropical Forestry and Forest Products (INTROP), Universiti Putra Malaysia, Malaysia
*Corresponding author, e-mail: imas.sitanggang@ipb.ac.id[1], razaliy@fsktm.upm.edu.my[2], norwati@fsktm.upm.edu.my[3], ainuddin@forr.upm.edu.my[4]

Abstract

Predicting hotspot occurrence as an indicator of forest and land fires is essential in developing an early warning system for fire prevention. This work applied a spatial decision tree algorithm on spatial data of forest fires. The algorithm is the improvement of the conventional decision tree algorithm in which the distance and topological relationships are included to grow up spatial decision trees. Spatial data consisted of a target layer and ten explanatory layers representing physical, weather, socio-economic and peatland characteristics in the study area Rokan Hilir District, Indonesia. Target objects were hotspots of 2008 and non-hotspot points. The result was a pruned spatial decision tree with 122 leaves and the accuracy of 71.66%. The spatial tree has produced higher accuracy than the non-spatial trees that were created using the ID3 and C4.5 algorithm. The ID3 decision tree had accuracy of 49.02% while the accuracy of C4.5 decision tree reached 65.24%.

Keywords: *spatial decision tree algorithm, spatial relationship, hotspot, forest fires, peatland*

1. Introduction

Predicting hotspots occurrence is considered as one of activities for fire prevention in order to reduce damages because of forest and land fires. Hotspots (active fires) indicate spatial distribution of forest and land fires. Hotspots occurrence models have been developed in several studies using geographical information systems and remote sensing technologies. In addition, data mining as one of growing areas in computer science has been applied to spatial forest fires datasets to obtain classification models for hotspots occurrence.

Decision tree is a famous method for classification tasks and it has been applied to a broad range of applications. Some of decision tree algorithms are ID3, C4.5 as a successor of ID3, and CART (Classification and Regression Tree). These algorithms are designed for non-spatial datasets. The different between spatial and non-spatial decision data is that in the spatial data, an object may have a significant influence on neighboring objects. Therefore, improvement of the non-spatial decision tree algorithm has been done by involving spatial relationships between two spatial objects.

Several studies have been conducted on spatial decision tree algorithms. The spatial decision tree algorithm was introduced in [1] based on the ID3 algorithm involving the spatial relationship Distance. The spatial binary tree algorithm was proposed in [2] that works on the dataset containing point, line, and polygon features. An extension of the CART method, called the SCART (Spatial Classification and Regression Trees), was developed in [3]. In the SCART, topological and distance relationships are used to test whether a predictive attribute belongs to the neighbor table. The SCART was applied to analyze traffic risk using accident information and thematic information about road networks, population census, buildings, and other geographic neighborhood details [3]. A spatial decision tree based on the ID3 algorithm that works on polygon features was introduced in [4]. The algorithm was applied to classify the average (per farm) market value of sold agricultural products based on climate, the distribution of the principal aquifers, crops cultivated, and the number of cattle and calves per area. The spatial entropy-based decision tree method was proposed in [5] which uses the spatial relation Distance to relate point and polygon features. The algorithm was used to classify gross values of agricultural output [5] and the air pollution index in main cities in China [6]. A new formula for

Received February 26, 2014; Revised April 28, 2014; Accepted May 15, 2014

spatial information gain was proposed in [7] by including spatial autocorrelation (neighborhood split autocorrelation ratio). The algorithm was applied to the raster format that is represented in a set of pixels.

This work developed a classifier for predicting hotspots occurrence using the spatial classification algorithm namely the spatial decision tree algorithm [8]. The algorithm is an extension of the coventional ID3 algorithm [9]. The new algorithm proposed in [8] can work on spatial datasets containing point, line and polygon features as representations of spatial objects. The formula of entropy and information gain in the ID3 algorithm were modified by involving two types of spatial relationships namely metric and topological to relate two spatial objects [8]. The spatial dataset used in this work contains forest and land fires data for the study area Rokan Hilir district in Riau Province Indonesia. In addition to physical, socio-economic, weather characteristics of the study area [8], this work includes peatland types and peatland depth to predict fires occurrence in peatlands.

A peatland fire is classified as a ground fire because the fire burn peat soil inside the peatlands and we can only see smoke visible on the surface. Therefore, peatland fires are not easy to handle compared to the fires in non-peatlands [10,11]. A study in [12] reports that Riau is one of provinces in Sumatra that has high deforestation because of forest fires. Riau province had about 4.044 million hectares (56.19 %) of peatland in 2002 and it made the province as the largest area of peatland in Sumatera Island and Kalimantan Island. For that, influencing factors for fire events in peatlands are considered in this study.

2. Research Method

2.1. Study Area and Forest Fires Data

This work developed the prediction model for hotspots occurrence based on the forest fires dataset for the study area Rokan Hilir district in Riau Province in Indonesia. Rokan Hilir is located in the area between 100°16' - 101°21' East Longitude and 1°14' - 2°30' North Latitude. It covers an area of 8,881.59 km^2 or about 10 percent of Riau's total land area [13].

The spatial forest fires data include physical, socio-economic, weather and peatland characteristics of the study area that may influence forest and land fire events. The data and its source are provided in Table 1.

Table 1. Data and its source

Data	Source
Spread and coordinates of hotspots 2008 (for creating models for hotspots occurrence prediction)	FIRMS MODIS Fire/Hotspot, NASA/University of Maryland
Spread and coordinates of hotspots 2010 (for model evaluation)	FIRMS MODIS Fire/Hotspot, NASA/University of Maryland
Weather data 2008 (in the NetCDF format): maximum daily temperature, daily rainfall, and speed of wind	Meteorological Climatological and Geophysical Agency (BMKG)
Digital maps for road, rivers, city centers, land cover, and administrative border	National Coordinating Agency for Survey and Mapping (BAKOSURTANAL)
Digital maps for peatland depth and peatland types	Wetland International
Inhabitant's income source	BPS-Statistics Indonesia

2.2. Spatial Relationship, Spatial Entropy and Spatial Information Gain

Spatial datasets for classification tasks are composed by some explanatory layers and one target layer. Each layer represents a set of spatial objects which is characterized by several spatial and non-spatial attributes. One of non-spatial attributes in an explanatory layer is the explanatory attribute that identifies objects in the layer. The target layer has a target attribute that stores class labels of the target objects.

All objects in a layer have a particular geometry type that may be either point, line or polygon. The geometry type of objects is presented in a spatial attribute of the layer. For instance, in this study the road layer represents a road network in which each road segment has the geometry type of line. Other layers in the dataset are the land cover layer and the target layer. Spatial objects in the land cover layer are polygon features, whereas objects in the target layer are point features indicating hotspots and non-hotspots.

Relation between spatial objects of two different layers is essential in spatial data mining systems. In our study, for example, hotspots occurrence in the target layer may be influenced by the existence of roads because roads open access for human to enter a forest and their activities may trigger forest fire events. Moreover, different land cover types may provide different risk levels of fires occurrence. For instance, fires are more likely take place in plantation areas than those are in settlement areas because farmers may use fires to open new plantations.

Spatial relationships allow us to include relations between two spatial objects in a dataset for a classification task. These relationships can be topological such as meet and overlap, as well as metric, for example distance. In spatial databases, a layer is represented as a relation and applying a spatial relation between two layers results a new relation. The structure Spatial Join Index (SJI) was introduced in [14] to implement spatial relationships in the relational database framework. The SJI is a new relation as the result of join index between two relations that consists of indices pairs each referencing a tuple of each relation. The pairs of indices refer to objects that meet the join criterion.

The concept of SJI was adopted in our previous work [8]. The work in [8] computed quantitative values resulted from topological and metric relationships. A topological relation between two spatial objects is calculated by performing the overlap operation. In addition to topological relationships, the algorithm involves a metric relationship namely distance from a spatial object to another spatial object. For example, applying the spatial relationship *overlap* on two polygons results an overlapping area with a certain extent. Moreover, we may also count how many hotspot points in a certain polygon or calculate distance between hotspot points to a nearest river segment. We denote these quantitative values, i.e. area, count and distance, as spatial measures of spatial relationships between two objects. Instead of using the SJI, our work proposes what we called Spatial Join Relation (SJR), as the result of a spatial relation between two layers [8]. The SJR contains spatial objects from the two layers and its associated spatial measures. The SJR of a new layer R is defined as follows [8]:

SJR = {(p, SpatMes(r), q | p in layer L_i, q in layer L_j, and r is a feature in R associated to p and q}. (1)

The spatial measure of a layer R, SpatMes(r), is used in the spatial entropy formula which replaces the number of tuples in a partition in the non-spatial entropy formula. The spatial entropy is defined as follows [8]. Let the target attribute C in the target layer S has *l* distinct classes (i.e. c_1, c_2, …, c_l), entropy for S represents the expected information needed to determine the class of tuples in the dataset and defined as

$$H(S) = -\sum_{i=1}^{l} \frac{\mathrm{SpatMes}(S_{c_i})}{\mathrm{SpatMes}(S)} \log_2 \frac{\mathrm{SpatMes}(S_{c_i})}{\mathrm{SpatMes}(S)} \tag{2}$$

SpatMes(S) represents the spatial measure of layer S that may be either area, count or distance.

The spatial decision tree algorithm partitions objects in the target layer S based on the explanatory (non-target) layer L. This step results a new layer L(v_j, S) for each possible value v_j in L. Each new layer is associated to a new partition. The expected entropy value for splitting is defined as follows:

$$H(S \mid L) = \sum_{j=1}^{q} \frac{\mathrm{SpatMes}(L(v_j, S))}{\mathrm{SpatMes}(S)} H(L(v_j, S)), \tag{3}$$

Spatial information gain for the layer L is given by the following formula.

$$Gain(L) = H(S) - H(S|L) \tag{4}$$

where *H*(*S*) and *H*(*S*|*L*) are given in Equation 2 and Equation 3 respectively. The layer L with the highest information gain, *Gain*(*L*), is selected as the splitting layer to partition the dataset.

2.3. Spatial ID3 Algorithm

The ID3 decision tree algorithm was developed by J. Ross Quinlan during the late 1970s and early 1980s. This algorithm has the principle that it builds the tree in greedy manner starting from the root, and selecting most informative features at each step [15]. The algorithm uses information gain to select the best feature at each step for splitting a dataset. Furthermore, the ID3 algorithm is designed for non-spatial datasets in which the input of the algorithm is a relation containing some objects of interest. All objects are characterized by several features. One of the features is a target feature that consists of class labels of objects, whereas other features are explanatory features that will be used to classify an object to a certain class label.

The ID3 algorithm has been improved in [4] such that the algorithm can be applied on a spatial dataset containing polygon features. On the other hand, spatial datasets may involve not only polygon features but also point and line features. Therefore in our previous work [8], we extended the ID3 algorithm based on several approaches in [4] so that the new algorithm can work on point, line and polygon features. Our proposed algorithm uses the spatial information gain provided in Equation 4 to select the best splitting layer from a set of explanatory layers.

Creating a spatial decision tree using the spatial decision tree algorithm [8] follows the basic learning process in the algorithm ID3 [9]. The algorithm works on spatial data stored in a spatial database. Before the algorithm is executed, the database contains only a set of explanatory layers and one target layer. When the algorithm works on the database, some new layers are produced as the result of spatial relations between two distinct layers. These new layers are created from existing explanatory layers, and the value v_j of predictive attribute in the best splitting layer. The value v_j is a selection criterion in the query to relate an explanatory layer and the best layer. Each new layer is associated with a set of tuples that relate objects in a layer to objects in another layer. This work considers this set of tuples as a smaller spatial dataset if one of two related layers is the target layer. Each tuple in the dataset has a spatial measure which is stored in the Spatial Join Relation (SJR). Inputs of the spatial ID3 algorithm are a spatial dataset, a set of explanatory layers, a target layer and a SJR. Output of the algorithm is a spatial decision tree. The tree has the same structure as that of the classical one in which the tree consists of a root node, internal nodes and leave nodes. The root node and internal nodes have the best splitting layers as its labels. Meanwhile, the labels of leave nodes are target classes of the target layer. There are some edges outgoing from the root node and internal nodes. The label of each edge is one of possible values in the best splitting layer.

2.4. Tree Pruning

Overfitting is one of issues that may be encountered when a decision tree algorithm is applied on real datasets. In this situation, as the decision tree become too large, the generalization error of decision tree starts to increase while its resubstitution error continues to decrease [16]. Resubstitution errors are misclassification errors on the training set, whereas generalization errors are misclassification errors on the testing set. Leaves in large trees may reflect noises or outliers that can increase generalization errors when the tree is applied on the testing set. One of methods to overcome overfitting is post-pruning in which the tree is fully grown at first, and then all subtrees of the tree at given nodes are pruned by removing its branches and replacing it with a leaf [17]. The new leaf is labeled with the majority class in the subtree.

3. Results and Discussion

3.1. Spatial Decision Tree for Hotspots Prediction

Applying the spatial ID3 algorithm on the forest fires dataset results a spatial decision tree which has 210 leaves. Accuracy of the tree on the training set is 76.51% meaning that 238 of 1013 target objects are incorrectly classified by the tree. Target objects are hotspots and non-hotspot points in the study area. Non-hotspot points were generated outside buffers of hotspots. The radius of a buffer for a hotspot is 0.907374 km. It was defined by processing burn areas extracted from the Landsat TM image. The first test layer of the tree is income source. This work prepared a testing set from the spatial database by applying several spatial operations. The testing set consists of 561 objects (235 positive examples and 326 negative examples). A positive example is an object with the true class, whereas a negative example is an object with the false class. Accuracy of the tree on the testing set is 71.12% meaning that 399 of 561 target objects are correctly classified by the tree.

The spatial decision tree as a prediction model for hotspots occurrence has the size of 613. Size of a tree is number of nodes including a root node, internal and leaves nodes. The number of classification rules generated from the tree is 134. A rule is obtained from a tree by creating a path from the root to a leaf. In order to obtain a simpler tree with the higher accuracy, the post-pruning method was applied to the tree. In this method, the tree is fully grown at first, and then all subtrees at given nodes are pruned by removing its branches and replacing it with a leaf [17]. This work implemented the post-pruning method up to 16 iterations. The last pruned tree has the accuracy of 71.66% and its size is 485. Starting from the second iteration, the highest accuracy of pruned trees for all iterations are the same i.e. 71.66%. However, the size of tree decreases from 599 in the second iteration to 485 in the 16th iteration. Therefore, the number of rules generated from the tree also declines. There are 108 rules generated from the simple pruned tree. Several rules are the following:

1. IF income_source = Plantation AND distance to the nearest road (m) ≤ 2500 AND 1500 < distance to the nearest river (m) ≤ 3000 THEN Hotspot Occurrence = True
2. IF income_source = Forestry AND land_cover = Bare_land AND 1 ≤ wind_speed (m/s) < 2 THEN Hotspot Occurrence = True
3. IF income_source = Forestry AND land_cover = Swamp THEN Hotspot Occurrence = TRUE
4. IF income_source = Forestry AND land_cover = Bare_land AND 0 ≤ wind_speed (m/s) < 1 AND 297 ≤ screen temperature (K) < 298 AND peatland_depth = D4 (Very deep/Very thick > 400 cm) THEN Hotspot Occurrence = False
5. IF income_source = Forestry AND land_cover = Paddy_field AND 0 ≤ wind_speed (m/s) < 1 THEN Hotspot Occurrence = False
6. IF income_source = Trading_restaurant THEN Hotspot Occurrence = False
7. IF income_source = Forestry AND land_cover = Mix_garden AND 0 ≤ wind_speed (m/s) ≤ 1 THEN Hotspot Occurrence = FALSE
8. IF income_source = Forestry AND land_cover = Plantation AND 0 ≤ wind_speed (m/s) ≤ 1 AND peatland_depth = Shallow/Thin (50-100 cm)THEN Hotspot Occurrence = FALSE
9. IF income_source = Forestry AND land_cover = Unirrigated_agri_field AND 2 ≤ precipitation (mm/day) ≤ 3 THEN Hotspot Occurrence = FALSE
10. IF income_source = Forestry AND land_cover = Paddy_field AND 0 ≤ wind_speed (m/s) ≤ 1 THEN Hotspot Occurrence = FALSE

3.2. Comparison between Spatial and Non-Spatial Classifiers

For comparison, the non-spatial decision tree algorithms namely C4.5 and ID3 have been applied on the forest fires dataset [18]. These algorithms are available in the data mining toolkit Weka 3.6.6. J48 is a module in Weka as Java implementation of the C4.5 algorithm. The accuracies of classifiers generated by these two algorithms were determined using the 10-folds cross validation method. In addition to non-spatial decision tree algorithms, a logistic regression model was calculated to predict hotspots occurrence [18]. Hotspots occurrence is considered as the dependent variable and determinant factors (environmental and human factors) influencing fire events are the independent variables. Table 2 summarizes the accuracy of the spatial and non-spatial classifiers as well as the number of rules generated from the trees.

Table 2. Accuracy of the classifiers and number of generated rules

Classifier	Accuracy	Number of generated rules
Spatial decision tree		
The Extended Spatial ID3 Decision Tree without pruning	71.12%	134
The Extended Spatial ID3 Decision Tree with pruning	71.66%	108
Non-spatial classifier		
ID3 Decision Tree	49.02%	270
C4.5 Decision Tree	65.24%	35
Logistic regression	68.63%	-

Table 2 shows that the proposed algorithm namely the spatial decision tree algorithm is superior among other methods i.e. non-spatial decision tree algorithms and logistic regression. The spatial ID3 without pruning performs well on the testing set with the accuracy of 71.12% compared to the classical ID3 (non-spatial ID3) with the accuracy of 49.02%. Furthermore, Table 2 shows that the spatial ID3 decision tree with pruning outperforms the C4.5 decision tree with 6.42% of accuracy higher than the C4.5 decision tree. Moreover, logistic regression has been used in several studies to determine the relation between hotspots occurrence and influencing factors of fire events. Applying this method to the forest fire dataset results the best regression model with the accuracy of 68.63% which is not better than the spatial decision tree algorithm that has the accuracy greater than 71%. According to these results, this work concludes that involving spatial relations in the decision tree algorithm produces the better classifiers for hotspots occurrence.

The spatial ID3 algorithm produces more simple trees compared to the ID3 algorithm. It can be inferred from the number of rules generated from the tree as shown in Table 2. The spatial ID3 algorithm without pruning gives 134 rules which is almost a half of the number of rules generated by the conventional ID3 decision tree i.e. 270. However, in term of the number of rules generated from the trees, the C4.5 algorithm outperforms the spatial ID3 algorithm with pruning where the C4.5 algorithm results only 35 rules and the proposed algorithm produces 108 rules (Table 2). The further study is required especially in the tree pruning method in order to obtain more simple spatial decision trees. On the other hand, the C4.5 decision tree has the accuracy of 65.24% that is slightly lower than the spatial ID3 decision tree with pruning which achieves the accuracy of 71.66%. Therefore, regardless the size of trees, the spatial ID3 algorithm with pruning has better performance than the C4.5 algorithm.

3.3. Tree Evaluation

The unpruned and pruned trees were applied to a new spatial dataset. The dataset contains the same explanatory layers as those for creating the tree and the FIRMS MODIS Fire/Hotspots in 2010. The number of hotspots in 2010 for Rokan Hilir area is 774. As many 726 points were randomly generated near any hotspot in 2010. To accomplish this task, buffers with the radius of 0.907374 km were created for each hotspot and then random points were generated outside the buffers. These random points are denoted as false alarm data. Along with hotspots in 2010 as true alarm data, false alarm data compose target objects in the new target layer.

A new dataset contains 707 objects (277 positive examples and 430 negative examples). Applying the spatial decision trees algorithm on the new dataset results the accuracy of 60.06% for the tree without pruning and 61.89% for the tree with pruning. Moreover, the tree is unable to classify some objects in the new dataset. There are 51 of 707 (7.21%) objects that cannot be classified by the tree without pruning. The number of unclassified objects decreases to 30 of 707 (4.24%) when the tree with pruning was executed on the new dataset. Table 3 gives characteristics of unclassified objetcs based on land cover, peatland type, peatland depth and income source. Most of unclassified objects are located in non-peatlands in which income sources of people living in these areas are mostly forestry and agriculture.

Table 3. Characteristics of unclassified objects

Explanatory attribute	True class	False class	Total
Land cover			
Plantation	2	6	8
Dryland_forest	2	7	9
Bare_land	0	1	1
Shrubs	2	2	4
Paddy_field	0	1	1
Swamp	3	0	3
Mix_garden	1	3	4
Peatland_type			
Hemists/Saprists(60/40),Moderate	2	4	6
Hemists/Saprists(60/40),Very_deep	2	2	4
Non_peatland	6	7	13
Saprists/min(90/10),Moderate	0	1	1
Saprists/min(50/50),Shallow	0	6	6
Peatland depth			
D1 (Shallow/Thin 50-100 cm)	0	9	9
D2 (Moderate 100-200 cm)	2	2	4
D3 (Deep/Thick 200-400 cm)	2	2	4
Non_peatland	6	7	13
income_source			
Other_agriculture	2	1	3
Forestry	0	10	10
Agriculture	8	9	17

4. Conclusion

This work applied the spatial ID3 algorithm on the spatial forest fires dataset. The dataset consists of physical, weather, socio-economic and peatland characteristics that may influence fires occurrence in the study area Rokan Hilir District, Indonesia. The result is a spatial decision tree for predicting hotspots occurrence with the accuracy of 76.51% on the training set and 71.12% on the testing set. Size of the tree is 613 and the number of rules generated from the tree is 134. To simplify the tree, the post-pruning method has been implemented. Applying this method on the spatial decision tree produces a pruned tree which is simpler than the unpruned tree. The pruned tree has the accuracy of 71.66% with income source as the first test layer. The size of the tree decreases to 485 and the number of generated rules declines to 108.

In comparison with the spatial ID3 algorithm, this work also applied the non-spatial decision tree algorithms i.e. ID3 and C4.5 on the forest fires dataset. The experimental results show that the proposed algorithm has better performance in term of accuracy than the two non-spatial algorithms. The accuracy of ID3 decision tree is 49.02% and the accuracy of C4.5 decision tree is 65.24%. Moreover, the spatial ID3 algorithm outperforms the logistic regression model that has the accuracy of 68.63%. The spatial ID3 algorithm has been tested to classify objects in the new forest fires dataset. The results show that there are 30 of 707 or about 4.24% objects which cannot be classified by the pruned tree. These unclassified objects mostly take place in non-peatlands in which income sources of people living in these areas are forestry and agriculture. Moreover, most of unclassified objects are located in plantation and dryland forest.

This work concludes that involving distance and topological relations between objects in the spatial classification task results the spatial decision tree as a model for predicting hotspots occurrence with the high accuracy.

Acknowledgment

This work was supported by Indonesia Directorate General of Higher Education (IDGHE), Ministry of National Education under Grant [number 1724.2/D4.4/2008]; and Southeast Asian Regional Center for Graduate Study and Research in Agriculture (SEARCA) under Grant [Ref. No. GCS10-2129].

References

[1] Ester M, Kriegel HP, Sander J. *Spatial Data Mining: A Database Approach*. Proceedings of the 5th International Symposium on Large Spatial Databases. Berlin. 1997: 47-66.

[2] Koperski K, Han J, Stefanovic N. *An Efficient Two-step Method for Classification of Spatial Data*. Proceedings of the International Symposium on Spatial Data Handling. Vancouver. 1998: 45-54.
[3] Chelghoum N, Zeitouni K, Boulmakoul A. *A Decision Tree for Multi-layered Spatial Data*. Proceedings of the Joint International Symposium on Geospatial Theory, Processing and Applications. Ottawa. 2002: 8-12.
[4] Rinzivillo S, Franco T. *Classification in Geographical Information Systems*. In: Boulicaut JF, Esposito F, Giannotti F, Pedreschi D. Editors. *Artificial Intelligence*. New York. Springer-Verlag. 2004: 374–385.
[5] Li X, Claramunt C. A Spatial Entropy-based Decision Tree for Classication of Geographical Information. *Transactions in GIS*. 2006; 10(3): 451-467.
[6] Zhao M, Li X. *An Application of Spatial Decision Tree for Classification of Air Pollution Index*. Proceedings of the 19th International Conference on Geoinformatics. Shanghai. 2011: 1-6.
[7] Jiang Z, Shekhar S, Mohan P, Knight J, Corcoran J. *Learning Spatial Decision Tree for Geographical Classification: A Summary of Results*. Proceedings of the 20th International Conference on Advances in Geographic Information Systems. California. 2012: 390-393.
[8] Sitanggang IS, Yaakob R, Mustapha N, Ainuddin AN. Classification Model for Hotspot Occurrences using Spatial Decision Tree Algorithm. *Journal of Computer Science*. 2013; 9(2): 244-251.
[9] Quinlan JR. Induction of Decision Trees. *Machine Learning*. 1986; 1(1): 81–106.
[10] Syaufina L, Nuruddin AA. Forest Fire in Peat Forest: An Overview. *Manajemen Hutan Tropika*. 2000; 6(1): 75–83.
[11] Adinugroho WC, Suryadiputra INN, Saharjo BH, Siboro L. *Manual for the Control of Fire in Peatlands and Peatland Forest. Climate Change, Forests and Peatlands in Indonesia Project*. Wetlands International, Indonesia Programme and Wildlife Habitat Canada. 2005.
[12] Wahyunto, Suryadiputra INN. *Peatland Distribution in Sumatra and Kalimantan-explanation of its Data Sets Including Source of Information, Accuracy, Data Constraints and Gaps*. Wetlands International, Indonesia Programme. 2008.
[13] Rokan Hilir District: Overview of district [Internet]. 2009. Riau: Rokan Hilir District; [cited 2012 May 30]. Available from: http://www.rohilkab.go.id/?tampil=linkandact=profilandid=4.
[14] Zeitouni K, Yeh L, Aufaure MA. *Join Indices as a Tool for Spatial Data Mining*. Proceedings of the International Workshop on Temporal, Spatial and SpatioTemporal Data Mining. Lyon. 2000: 102-114.
[15] Marsland S. *Machine Learning: An Algorithmic Perspective*. Boca Raton. CRC Press. 2009: 133-139.
[16] Tan P, Steinbach M, Kumar V. *Introduction to Data Mining*. Boston. Pearson Addison Wesley. 2006: 172-176.
[17] Han J, Kamber M. *Data Mining: Concepts and Techniques*. Second Edition. San Francisco. Morgan Kaufmann. 2006: 304-306
[18] Sitanggang IS, Yaakob R, Mustapha N, Ainuddin AN. Predictive Models for Hotspots Occurrence using Decision Tree Algorithms and Logistic Regression. *Journal of Applied Sciences*. 2013; 13(2): 252-261.