Available online at www.sciencedirect.com

ScienceDirect

Procedia Environmental Sciences 33 (2016) 317 – 323

Procedia Environmental Sciences

The 2nd International Symposium on LAPAN-IPB Satellite for Food Security and Environmental Monitoring 2015, LISAT-FSEM 2015

Web-Based clustering application using Shiny framework and DBSCAN algorithm for hotspots data in peatland in Sumatra

Rachma Hermawati* and Imas Sukaesih Sitanggang

Department of Computer sciences, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, Indonesia

Abstract

Forest and land fires currently have become serious problems in Indonesia. Peatlands are frequently burnt because of its character istic, i.e. combustible when it was in dried condition. In the previous work, hotspots as on indicators for forest and land fires inclu ding in peatland were analyzed by applying density based clustering algorithm namely Density-Based Spatial Clustering Algorith m with Noise (DBSCAN). Clustering results hotspots distribution that can be used for preventing and controlling fire events. Thi s research aims to build a web-based clustering application for grouping hotspot data in peatlands in Sumatra using the Shiny fra mework, that is available in programming language R. Clustering was performed on hotspots data in peatland in 2002 and 2013 u sing the DBSCAN algorithm. This algorithm finds clusters by identifying areas that have a high hotspots density. The application was successfully built and has several features, namely: a) clustering hotspots, b) visualization of clustering results based on a ty pe of landuse, land depth, and peat type, c) providing the value of within cluster for cluster evaluation, and d) displaying a summ ary of clustering results. These features have been tested using the blackbox approach and the test results show that the features w ork properly and produce outputs in corresponding to the test scenario.

Peer-review under responsibility of the organizing committee of LISAT-FSEM2015

Keywords: Clustering; DBSCAN; hotspots; peatland fires; R; Shiny

* Corresponding author. Tel.: +62-857-1525-7152
E-mail address: rachmahermawati@gmail.com

Peer-review under responsibility of the organizing committee of LISAT-FSEM2015
doi:10.1016/j.proenv.2016.03.082

1. Introduction

Peatland area in Indonesia is estimated at 12.8 million hectares or about 10.8 percent of the land area of Indonesia [1]. The total area of about 4.5 million acres or 35 percent are found in Sumatra. At this moment, a peat thickness had shrinkage caused by forest-and-land fires [1]. Forest-and-land fires disaster is a serious problem that must be faced by Indonesia almost every year during the dry season. Fires that have occurred not only on dry land but also on wetlands (particularly peat). Peatland fires are much more difficult to be handled than the fires that have occurred in the soil minerals forest or the highlands [2] because a fire in peatland is difficult to know its spread. The fire could spread to the large area or extends to a location much further in which the fire cannot be seen from the surface. In addition, during the dry season, the surface of the peatland will be drying fast and can turn into charcoal-like nature that causes flammable peatland [2]. The incident of forest and peatland fires that recently happened in Indonesia left a very great damage. Since February 20 through March 11, 2014, Global Forest Watch detected 3101 high confidence hotspots alerts in Sumatra island using NASA's Active Fire Data [3]. The Fig. exceeds the total number of 2643 hotspots alerts detected on March 13 to June 30, 2013, at the peak of the previous fires and haze crisis [3].

Adinugroho et al. [2] informs that data hotspots can be used as one indicator of a possible occurrence of fires. Therefore, it is necessary to do hotspots analysis to be used as reference in the control of forest-and-land fires. One of the hotspots data analysis techniques that can be used is the clustering technique. Clustering algorithms have been applied on hotspots data in the previous works. The clustering method k-means was used in [4] to identify global and collective outliers on hotspot frequency in the study area Riau Province, Indonesia. In addition, the work [5] developed a web-based Online Analytical Processing (OLAP) to visualize k-means clustering results on hotspot data in Indonesia. Another algorithm for clustering is the Density-Based Spatial Clustering Algorithm with Noise (DBSCAN) that is designed to find clusters and noises on spatial data [6].

DBSCAN is clustering algorithms that developed based on the level of density of the data. DBSCAN looking the center of objects, that is the objects that have a dense neighborhoods. DBSCAN connects the center of the object with its neighborhoods to form 1 areas as cluster [7]. DBSCAN algorithm first introduced by Esther *et al.* [6] in 1996 designed to look for clusters and noises on spatial data. This algorithm has two main parameters, namely Eps and MinPts. Eps is the minimum distance from the center of the cluster object in neighborhood, while MinPts is the minimum amount of neighbor from the central object of a cluster [6].

A previous research conducted clustering on hotspots data of Sumatra using the DBSCAN algorithm [8], but it had not yet built a web-based clustering application for hotspots data. A web-based clustering application is needed to make easier to analyze clustering results of hotspots data because it can be accessed from anywhere and anytime. Many frameworks that can be used to build a web-based application, one of which is Shiny. Shiny is a framework that uses the programming language R. Shiny is easy to use to build interactive web applications. Automatic binding reaction between input, output, and pre-built widgets causes Shiny to be able to build responsive applications. Therefore, this research utilizes the Shiny framework in programming language R to develop a web-based application for clustering hotspots data using the DBSCAN algorithm.

2. Methods

The data used in this research are hotspots data in 2002 and 2013 in peatland in Sumatra. Data were obtained from the previous research [8] that sourced from the Ministry of Forestry of the Republic of Indonesia and have been through the preprocessing data stage that are data cleaning and data reduction. Data cleaning and data reduction was aim to get the attribute longitude and latitude. There are 13253 hotspots of Sumatra peatland in 2002, and 3335 hotspots in 2013. The attribute used on both data are longitude and latitude. These data are stored in a file with csv format.

In addition to hotspots data in peatland in 2002 and 2013, spatial data in shapefile are used for visualization the clustering results. Shapefile data used are the shapefile of province borders in Sumatra in the file format -.shp, .dbf, .shx, .sbn, and .spx and the shapefile of peatland in Sumatra in the file format .shp, .dbf, .shx, .prj, and .qpj. Both these shapefile data were obtained from previous research [8] that sourced from Wetlands International Program Indonesia.