IOP Conf. Series: Earth and Environmental Science **581** (2020) 012045 doi:10.1088/1755-1315/581/1/012045

Gradient Boosting Tree for Land Use Change Detection Using Landsat 7 and 8 Imageries: A Case Study of Bogor Area as Water Buffer Zone of Jakarta

Juni Handoko[1], D E Herwindiati[1], and J Hendryli[1]

1 Faculty of Information Technology, Universitas Tarumanagara, Jakarta, Indonesia
E-mail: junihandoko97@gmail.com

Abstract. The land use change in Bogor regency need to be studied, since it acts as a water buffer zone for the sorrounding area, which includes the capital city of Jakarta. This study aims to analyze the land use changes in Bogor Regency using the gradient boosting tree model. Landsat 7 and 8 imageries of Bogor area in 2008, 2011, 2014, and 2017 are used as the case study. The images are cropped into sub-district level and classified into four classes, which are green area, partial green area, impervious land, and partial impervious land. By comparing two images of classification result between two different years, the land use changes can be determined. This study shows that most land use changes from 2008 to 2017 occur in Sukamakmur sub-district with a percentage of 69.31% (134.0757 km^2). Based on the type of land, most land use changes are from impervious area to green area.

1. Introduction

Bogor regency is historically categorized as water buffer zone, because it is strategically located and has many open spaces and forests. The location, which is strategically sorrounded by several mountains, such as Pancar Mountain, Salak Mountain, Gede Mountain, and Pangrango Mountain, serves as provider of water sources for daily needs, such as drinking water and sanitation. Furthermore, Puncak area of Bogor regency has also been assigned as a special area for soil and water conservation, tourism, and also as the buffer area of the capital city of Jakarta [1]. But lately, many damages in the water buffer zone have caused Bogor regency to be no longer considered as water buffer zone for Jakarta. One of the main factors that caused this is the rise of residential development, especially the construction of villas at the Puncak area, such that the area can not absorb water maximally.

Solutions are needed to overcome the imbalance of green and impervious area, such as controlling the development of settlements and tourist villas in the Puncak area and also adding more green open spaces in the Bogor regency. The execution of these solutions needs regional, which is usually done manually and takes time. Hence, another method for monitoring is needed. Remote sensing system can be utilized for this matter. The data can come from satellite imagery, such as from Landsat satellites. This purpose of this study is to detect and analyze the land use changes in Bogor Regency from 2008 to 2017 from Landsat 7 and 8 satellite imagery.

The classification method employed for this study is the gradient boosting tree. The model has been applied successfully in many cases. For example, it is frequently used in the anomaly detection system where the data are often highly imbalanced, such as DNA sequences [2], credit card transactions [3], or cyber security [4], the gradient boosting tree can achieve good performance. This model is suitable for this study because the impervious and green area in Bogor Regency are also imbalanced. Gradient boosting tree performs the optimization with function space so the custom loss functions can be used much easier. Furthermore, the boosting in this method handles the unbalanced datasets by strengthening the impact of the positive class [5].

The gradient boosting tree is one of ensemble learning methods that uses the principle of boosting where many weak learners are built for creating a strong model. The main idea of this algorithm is to construct new base-learners to be maximally correlated with negative gradient of loss function. For a better intuition, rather than applied arbitrary, if the loss function is classic squared-error loss, the learning procedure would result in consecutive error-fitting [6]. Gradient boosting tree can use, for example, a classification and regression tree (CART) as the weak learner for building a model that usually called the strong learner. That model will be used for classifying Bogor regency's area into four class of land types. The land types are the green area, partial green area, impervious land, and partial impervious land.

This study uses Landsat 7 and Landsat 8 remote sensing data of Bogor regency as the case study. The spectral bands from Landsat 7 and Landsat 8 that is used are blue, green, red, near infrared, and both short wave infrared channels. Before the classification step, the Landsat 7 and 8 imageries need to be pre-processed first. The pre-processing step consists of three processes, which are the radiometric correction, transformation, and image cropping. Furthermore, in 2003, the SLC scan line corrector instrument of Landsat 7 failed permanently causes large gaps at the edges of the image. Therefore, a gap filling procedure have to be applied to restore the image [7]. After the gap filling procedure, radiometric correction is employed to reduce or correct errors in the digital numbers of the images. Radiometric correction method that is used is the dark subtraction method to remove the effects of atmospheric scattering from an image. The second pre-processing process is the transformation of the Landsat 8 imagery. This process transforms the digital number of Landsat 8 image, such that it has digital numbers in the interval of 0 to 256, similar to the Landsat 7 images. The final pre-processing step is the cropping of Landsat 7 and 8 imageries of Bogor regency into 40 sub-districts. After the pre-processing, the images can be classified to determine the land type and land use changes in the Bogor regency.

2. Landsat 7 and 8 imageries

In 1972, Landsat earth observation was introduced for monitoring land cover and land use globally. Landsat 7 was launched on April 15, 1999 and bring the new sensor named Enhanced Thematic Mapper (ETM+) sensor. Landsat 7 mission is still the same as previous Landsat, which is the observations and also demonstrates significant progress in precise numerical radiometry, spectral differentiation, and seasonally repetitive monitoring [8]. On May 31, 2003, the Scan Line Corrector (SLC) instrument on Landsat 7 is damaged and delivers data with missing gaps. The imaged area is duplicated with a width that increases toward the edge of the scene. As the result, when the duplicated areas are removed, the images are leaved with gaps in the form of black lines. See figure 1

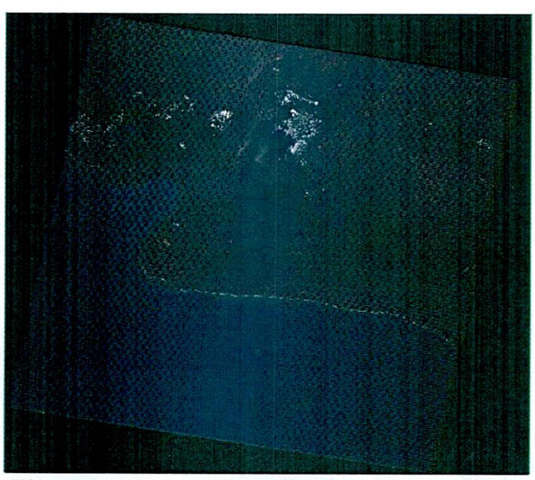
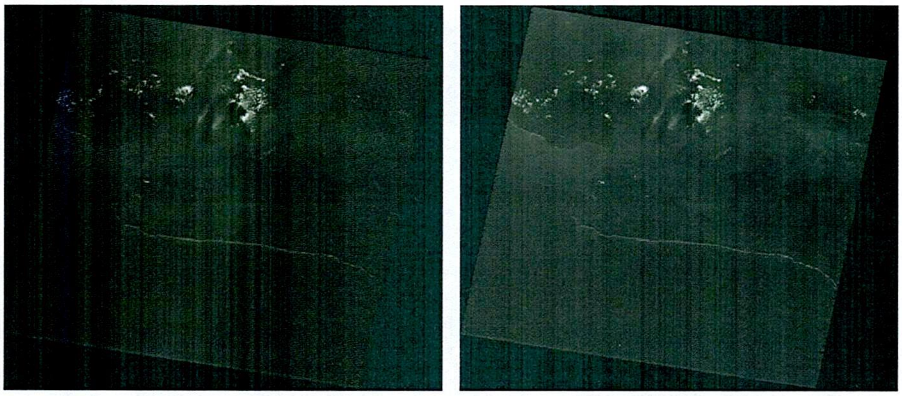


Figure 1. Example of Landsat 7 SLC-off scene.

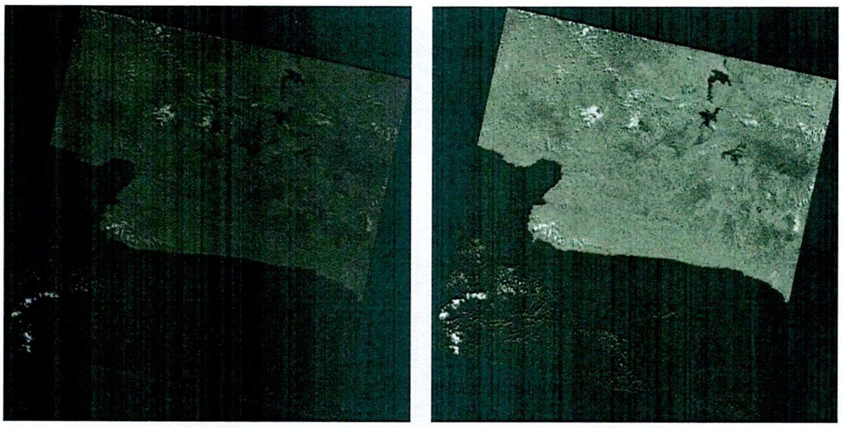
Landsat 8 was launched on February 11, 2013 and bring two new sensors that named Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). The OLI sensor consists of 9 bands with a spatial resolution of 30 meters for bands 1, 2, 3, 4, 5, 6, 7, 9 and the TIRS consists of 2 thermal bands.OLI sensor also introduces two new bands which is deep blue coastal band that is used for coastal areas detection and shortwave-infrared cirrus band for Cirrus cloud detection.

In this study, the blue, green, red, near infrared, and both short wave infrared channels from Landsat 7 and 8 are used. The images are pre-processed with the radiometric correction, transformation, and cropping. For Landsat 7 imagery, the gap filling process and dark subtraction method to remove the effects of atmospheric scattering are also employed. In contrast, for the Landsat 8 imagery only the dark subtraction method is applied. Moreover, the transformation process is used to transform Landsat 8's 16-bit value to 8-bit so that it has similar value range to Landsat 7 images. See figure 2 and 3



(a) (b)

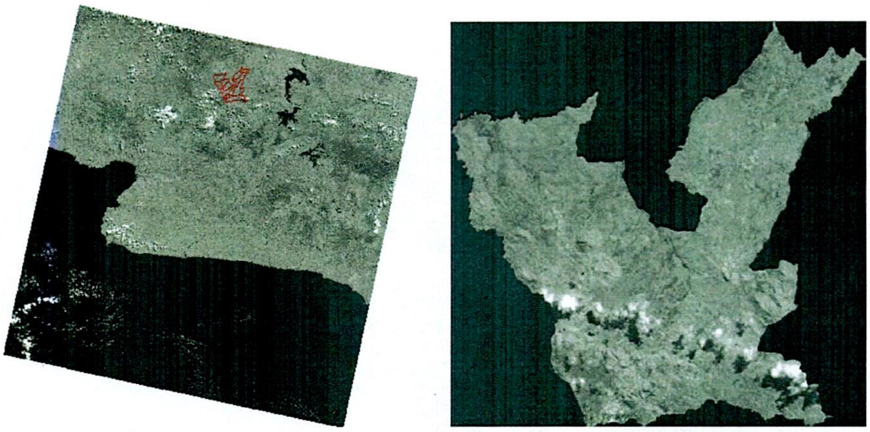
Figure 2. Landsat 7 imagery (a) before and (b) after applying the gap filling process and dark subtraction method.



(a) (b)

Figure 3. Landsat 8 imagery (a) before transformation and (b) after transformation.

The last process is cropping Landsat 7 and Landsat 8 imageries of Bogor regency. The cropped images are used in the training and testing steps as the input data. For training the model, the shapefiles of green area, partial green area, impervious land, and partial impervious land are created first. Landsat 7 and Landsat 8 imageries are cropped into certain shape according to the used shapefile. For testing process, the shapefiles consist of 40 shapes of Bogor sub-district, so that the mapping of land use changes can be more detailed. See figure 4



(a) (b)

Figure 4. Cropping result of (a) Landsat 8 imagery to (b) sub-district of Bogor regency.

3. Gradient boosting tree method

Gradient boosting tree uses classification and regression tree (CART) as the weak learner for building $F_m(x)$ model as strong learner. This method using forward stagewise additive modelling principle that constructs new base-learners to be maximally correlated with the negative gradient of the loss function. For classification problems, K least squares tree are constructed at each iteration and the loss function that is used is the multinomial deviance. Each tree is fit to its respective negative gradient vector g_{km}[9].

$$-g_{ikm} = I(y_i = g_k) - p_k(x_i) \tag{1}$$

Here is the algorithm of gradient boosting tree for K-class classification [6]:

1. Initialize $f_{k0}(x) = 0, k = 1,2, \dots, K$.
2. For $m = 1 \ to \ M$:
   (a) Set $p_k(x) = \frac{e^{f_k(x)}}{\sum_{l=1}^{K} e^{f_l(x)}}, k = 1,2, \dots, K$.
   (b) For $k = 1 \ to \ K$:
      i. Compute $r_{ikm} = p_k(x_i), i = 1,2, \dots, N$.
      ii. Fit a regression tree to the targets $r_{ikm}, i = 1,2, \dots, N$, giving terminal regions R_{jkm}, $j = 1,2, \dots, J_m$.
      iii. Compute $\frac{K-1}{K} \frac{\sum_{x_i \in R_{jkm}} r_{ikm}}{\sum_{x_i \in R_{jkm}} |r_{ikm}|(1-|r_{ikm}|)}, j = 1,2, \dots, J_m$.
      iv. Update $f_{km}(x) = f_{k,m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jkm} I(x \in R_{jkm})$.
3. Output $f_{k0}(x) = f_{kM}(x), k = 1,2, \dots, K$.

The first step is initialize the base learner $f_0(x)$. After that, proceeding to second step. On m-th iteration, decision tree as the weak learner is built to improve the $F_{m-1}(x)$ without changing the parameters of previous model. At each iteration, a loss function is calculated and produces a residual value in the form of a negative gradient. In the next iteration, residual value is minimized by building the next tree. The result of this process is strong learner as new model that will be used to do the classification process.

4. Training process

4.1. Data Input

The input for the training are some images from sub-district of Bogor regency's that are cropped from Landsat 7 and 8 imageries. They consist of six attributes or features and labeled to four land types. The attributes are the spectral bands of Landsat 7 and 8 imageries, which are the blue, green, red, near infrared, and two short wave infrared channels. Every pixels of the images are labeled with one of the four land types. There are 22709 pixels of green area, 22770 pixels of partial green area, 23136 pixels of impervious land, and 25066 pixels of partial impervious land. Table 1 gives the sample of the training dataset.

Table 1. Examples of training data.

Band Blue	Band Green	Band Red	Band NIR	Band SWIR-1	Band SWIR-2	Land Type
76	60	53	61	84	50	Green
77	59	53	62	86	49	Green
76	60	55	65	84	49	Green
76	60	61	60	84	51	Green
80	62	59	66	96	60	Green

The result of the training process is a gradient boosting tree model for classifying the land type. The model is tested using test data that are also cropped from Landsat 7 and 8 imageries. The test data consist 6862 pixels of green area, 7449 pixels of partial green area, 7352 pixels of impervious land, and 7136 pixels of partial impervious land.

4.2. The model

The model of gradient boosting tree method for training process is a decision tree. The decision tree from the last iteration in the training process is used as the model for classification. In the decision tree model, there area some variables, such as band and MSE (mean squared error) which denotes the features and the loss function value, respectively. Both variables are used to determine the split that forms the parent and terminal node of the tree. Figure 5 shows the tree obtained from the training process.

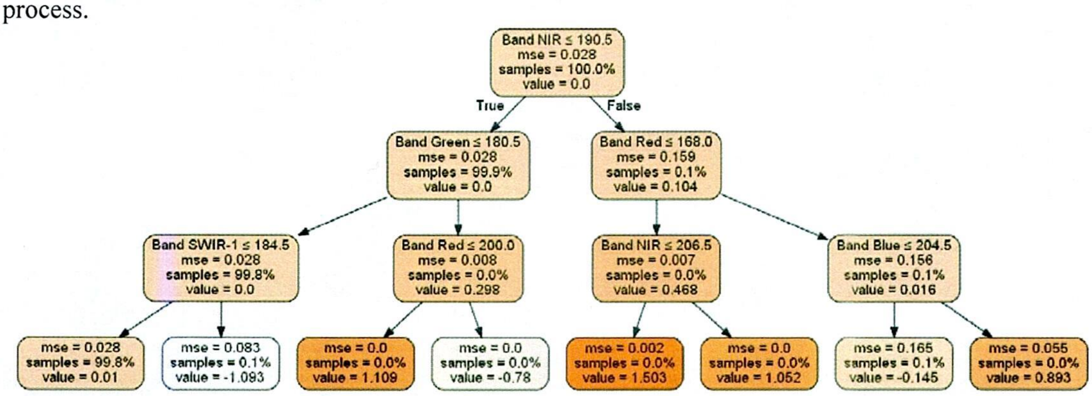


Figure 5. The model of gradient boosting tree.

4.3. Model evaluation

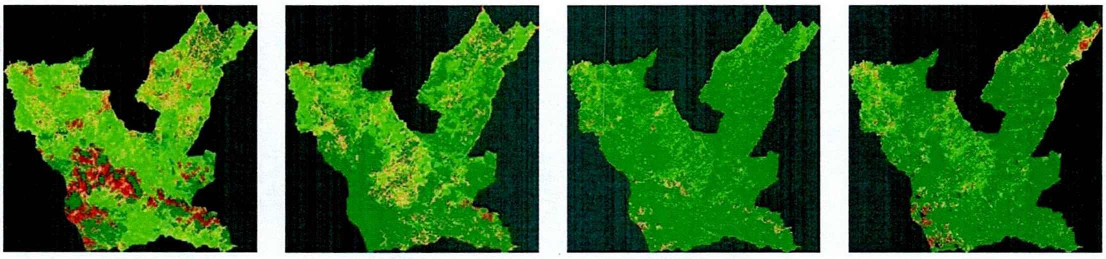
The experiment is conducted five times. On each experiments, the number of estimators which denotes the number of weak learners in the gradient boosting tree, is tuned. The best model is chosen according to the best F1-score. From five experiments (see Table 2), the model on the third experiment with 90 estimators performs best. The F1-score is 69.04%. Furthermore, the training duration is also considered. The best model can be trained fairly quick too.

Table 2. Experiment of training process

No	Number of Estimators	Training Loss	F1-score (%)	Duration (secs)
1	100	44764.4430	68.85	32.48
2	80	46873.4212	68.77	20.99
3	90	45759.9793	69.04	23.64

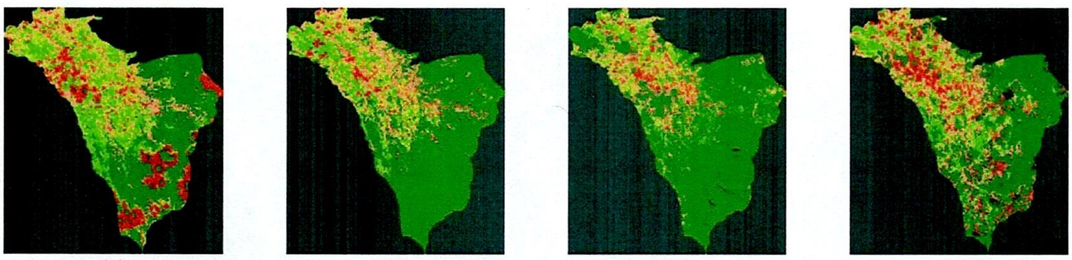
4	200	39218.0638	68.98	51.80
5	300	36490.5757	68.56	85.00

5. The mapping of bogor regency
Using Landsat 7 and Landsat 8 imageries of sub-districts in Bogor regency from 2008, 2011, 2014, and 2017, Figure 6 and 7 shows the classification result of sub-district Sukamakmur and Cisarua. The four land types are described by dark green which represents the green area, light green color for partial green area, dark red color for the impervious lands, and light red color to represent the partial impervious lands. It should be noted that the spatial resolution of the Landsat images is 30 meters.



(a) (b) (c) (d)

Figure 6. Classification result of Sukamakmur sub-district from (a) 2008, (b) 2011, (c) 2014, (d) 2017.

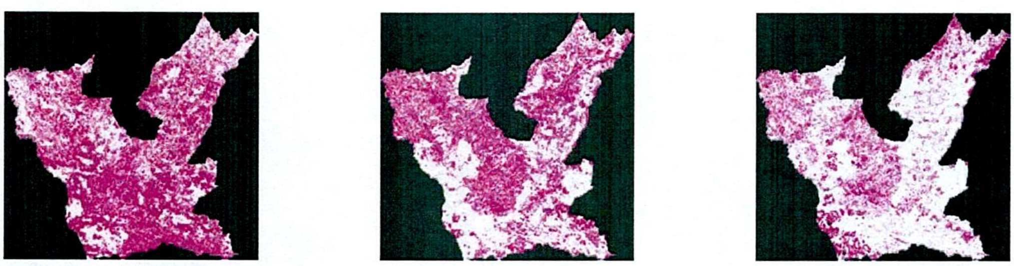


(a) (b) (c) (d)

Figure 7. Classification result of Cisarua sub-district from (a) 2008, (b) 2011, (c) 2014, (d) 2017.

6. Land use change

6.1. Land use change from impervious land to green area
In this section, we show the land use change from impervious land to green area observed from 2008, 2011, 2014, and 2017. Figure 8 shows the example of land use change mapping of Sukamakmur sub-district. The purple color represents the change, while white color represents area with no land use change.



(a) (b) (c)

Figure 8. Land use change from impervious to green area of Sukamakmur sub-district from (a) 2008 to 2011, (b) 2011 to 2014, and (c) 2014 to 2017

Table 3. Total area of Sukamakmur sub-district for each land types

Year	Green Area (km^2)	Partial Green Area (km^2)	Impervious Land (km^2)	Partial Impervious Land (km^2)
2008	36.5733	101.1006	18.5724	37.2960
2011	101.9313	58.1409	1.6839	31.7862
2014	155.6496	31.1274	0.3708	6.3837
2017	145.9296	31.7286	3.2076	12.1311

Table 4. Total area of land use change from impervious land to green area in Sukamakmur sub-district

Type of land use change	Year From	Year To	Total area (km^2)	Percentage of land use change (%)
From impervious land to green area	2008	2011	10.5102	5.43
	2011	2014	0.7929	0.41
	2014	2017	0.1251	0.06
From partial impervious land to green area	2008	2011	18.3231	9.47
	2011	2014	22.1274	11.43
	2014	2017	2.9601	1.53

Table 3 shows the total area of every land types in Sukamakmur sub-district in 2008, 2011, 2014, and 2017. Table 4 describes the total area of land use change from impervious land to green area for each years. While total land area of Sukamakmur sub-district is 193.4307 km^2, the total size of land use change from impervious land to green area is 11.4282 km^2 and land use change from partial impervious land to green area is 43.4106 km^2.

6.2. Land use change from green area to impervious land

Figure 9 shows the result of land use change at Cisarua sub-district which is the area with the highest percentage of land use change from green area to impervious land.

(a) (b) (c)

Figure 9. Land use change on Cisarua sub-district from (a) 2008 to 2011, (b) 2011 to 2014, (c) 2014 to 2017

Table 5. Total area of Cisarua sub-district for each land types.

Year	Green Area (km^2)	Partial Green Area (km^2)	Impervious Land (km^2)	Partial Impervious Land (km^2)
2008	26.1585	20.1987	9.7515	16.2108
2011	42.1119	11.9565	2.7045	15.5466
2014	48.0114	9.4608	2.4822	12.1707
2017	32.6241	12.6873	8.6130	17.5284

Table 6. Total area of land use change from green area to impervious land in Cisarua sub-district.

Type of land use change	Year		Total area (km^2)	Percentage of land use change (%)
	From	To		
From green area to impervious land	2008	2011	0.1161	0.16
	2011	2014	0.0540	0.07
	2014	2017	3.2148	4.47
From partial green area to impervious land	2008	2011	0.5202	0.72
	2011	2014	0.4140	0.57
	2014	2017	0.9432	1.31

Table 5 shows the total area of each land types in Cisarua sub-district on 2008, 2011, 2014, and 2017. Furthermore, table 6 shows the total and percentage of land use change from green area to impervious land from 2008 to 2011, 2011 to 2014, and 2014 to 2017. It should be noted that the total area of Cisarua sub-district is 72.3195 km^2. The total of land use change from green area to impervious land is 3.3849 km^2 and land use change from partial green area to impervious land is 1.8774 km^2.

7. References

[1] Lisnawati Y and Wibowo A 2009 Analysis of Land Carrying Capacity in Puncak Area, Bogor District *Jurnal Penelitian Hutan Tanaman* **6** 45-54
[2] Sahakyan A B, Chambers V S, Marsico G, Santner T, Antonio M D, and Balasubramanian S 2017 Machine learning model for sequence-driven DNA G-quadruplex formation *Scientific reports* **7** 1-11
[3] Rushin G, Stancil C, Sun M, Adams S, Beling P 2017 Horse race analysis in credit card fraud—deep learning, logistic regression, and Gradient Boosted Tree *Systems and Information Engineering Design Symposium (SIEDS)* pp 117-121
[4] Bansal A and Kaur S 2018 Extreme Gradient Boosting Based Tuning for Classification in Intrusion Detection Systems *International Conference on Advances in Computing and Data Sciences* pp 372-380
[5] Frery J, Habrard A, Sebban M, Caelen O and He-Guelton L 2017 Efficient top rank optimization with gradient boosting for supervised anomaly detection *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* pp 20-35
[6] Natekin A and Knoll A 2013 Gradient boosting machines, a tutorial *Frontiers in neurobotics* pp 7-21
[7] Herwindiati D E, Djauhari M A, and Jaupi L 2013 Robust Classification of Remote Sensing Data for Green Space Analysis *Journal of Mathematics and System Science* **3** 180-186
[8] Goward S N, Masek J G, Williams D L, Irons J R, and Thompson R J 2001 The Landsat 7 mission: Terrestrial research and applications for the 21st century *Remote Sensing of Environment* **78** 3-12
[9] Hastie T, Tibshirani R and Friedman J 2009 *The elements of statistical learning* (New York: Springer-Verlag) pp 359-387