Land Use Change Using Least Absolute Shrinkage and Selection Operator Regression in Jakarta's Buffer Cities

Devin Budi[1], Dyah Erny Herwindiati[2], Janson Hendryli[3]

[1,2,3]Faculty of Information Technology, Universitas Tarumanagara, Jakarta, Indonesia
Email: [1]devin.535170003@stu.untar.ac.id, [2]dyahh@fti.untar.ac.id, [3]jansonh@fti.untar.ac.id

Abstract—Land use change is one of the problems that need to be considered in city development. As buffer cities for Jakarta, the Bekasi, Bogor, Depok, and Tangerang areas need to observe the land use change by its subdistrict as a result of residential and industrial development for the city of Jakarta. The goverments need to watch the land use change to impervious area by subdistrict governments need to watch for land-use change to impervious land by sub-district to maintain the minimum standard of non-impervious land. In machine learning, regression could be used to classify land into impervious or non-impervious land. One of the examples of regression is Least Absolute Shrinkage and Selection Operator (LASSO). This study aims to rank a land use change by subdistrict area in Jakarta's buffer cities using Landsat 7 and Landsat 8 imageries. As a result, the highest land use change from 2017 to 2020 occurred in Bogor City (27.03% of its area) while the highest land use change to impervious in subdistrict happened in Tajur Halang subdistrict (50.69% of its area)

Keywords— *Classification, Land Use Change, Landsat Imagery, Radiometric Correction, Remote Sensing*

I. INTRODUCTION

Impervious surface is a surface area that are covered by water-resistant materials such as concrete, iron, glass, and plastic while non-impervious surface is a surface area that have not undergone land cover substance and it can absorb water such as forest, plantation, farm, green open space, savanna, and garden. [1] According to Badan Standarisasi Nasional, impervious area is an area that has undergone a natural or semi-natural of land cover substance which is an artificial cover that is usually waterproof and relatively permanent.[2] As the capital city, Jakarta carries out all political, industrial, economic, and social activities so it makes people want to live in Jakarta. Jakarta also undertakes the city development by taking the green surface for the industrial and residential areas. It also affects the surrounding areas such as Bekasi, Bogor, Depok, and Tangerang.

Urbanization is the process through which cities grow, and higher percentages of the population comes to live in the city.[3] According to Badan Pusat Statistik in Indonesian Population Census, in 2005, around 8,839,247 people lived in Jakarta. In 2010 Jakarta has 9,607,787 and in 2015 it has 10,154,134 people.[4][5][6] The population continues to increase from year to year due to urbanization. As a result, Jakarta cannot accommodate for urbanization people and Jakarta needs to expand city development to surrounding area. Bekasi, Bogor, Depok, and Tangerang are areas around Jakarta that are used as Jakarta's

expanding area or known as buffer cities. Land use change is an increase in land use from one use to another in accordance with a reduction in the type of land use over time, or a change in land use within a certain period of time.[7] Jakarta's city development turned non impervious area in its satelite cities into industrial and resident areas or known as impervious area. Population density causes several problems, including food availability, green area availability, facilities and infrastructure availability, and environmental damage. Environmental damage that has occurred includes landslides, floods, clean water crises, drought and global elimination.[8] Based on Law Number 26 Year 2007 Article 29 concerning Spatial Planning, Cities are required to provide and use Green Open Space with an area proportion of at least 30% of the city's area.[9] The purpose of this study is to find out the proportion of impervious land within every district in Jakarta's buffer cities and rank the impervious land in the regency/city. The data can be use for every district government to evaluate their urban planning or improving green space area and for city/regency government to coordinate within district to solve the ideal urban planning.

In this paper, the land is classified into three classes, namely green land, partial green land, and impervious land based on normalized difference vegetation index. After that analyzing the impervious land among the district and sub-district areas. Also analyzing the land use change from one use to another within a certain period of time in district and sub-district areas. After analyzing the impervious land. This paper also ranking the land use change from green or partial green area to impervious area and also analyzing that have the more impervious land.

II. Methods

A. Least Absolute Shrinkage and Selection Operator (LASSO) Regression

LASSO is a linear regression method by shrinking the regression coefficients of the predictor variables whose values are displayed with errors. This particular type of regression is well-suited for models showing high levels of muticollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.[10] The existence of outliers in multiple linear regression is a problem because outliers can cause the formation of a regression parameter model to be less accurate. In overcoming this problem, statisticians try to find alternative estimates of other parameters that are better

at overcoming outliers. One of the methods suggested in multiple linear regression is Least Absolute Shrinkage and Selection Operator with Least Angle Regression (LASSO LARS). The LASSO estimator is defined as follows:

$$\beta^{LASSO} = \arg\min\left\{\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j}^{k}\beta_j X_{i,j}\right)^2\right\} + \lambda\sum_{j=1}^{p}|\beta j|$$

Where:
$Y = dependent\ vector$
$X = independent\ variable$
$\beta = constant$
$n = number\ of\ observations$
$p = number\ of\ independent\ variabel$
$\lambda = the\ amount\ of\ shrinkage.$

In general, the algorithm for applying the LASSO LARS Regression can be summarized into the steps below:
1. First the data centering and rescalling process is carried out, then look for a vector that is proportional to the correlation vector between the predictor variables with the error.
$$\hat{C} = X^T(y - \hat{\mu})$$
2. Determine the greatest absolute moment correlation with the following equation
$$\hat{C} = \max\{|\hat{c}_j|\}$$
3. Determine X_A with set A as the set of active indices that belong to the predictor variable {1, 2, 3, ..., m} which is determined from the largest absolute correlation value with the formula:
$$X_A = \left\{\ldots s_j X_j^* \ldots\right\}; j \in A$$
4. Calculating the value of the equiangular vector. the equiangular vector is a vector that divides the angles of the XA columns into equal sizes with angles less than 90°. The equiangular vector value can be found using a formula
$$u_a = X_A\omega_A \text{ while } \omega_A = A_A G_A^{-1} 1_A$$
5. Calculate the product vector
$$a = X'u_A$$
6. Calculating $\hat{\mu}_A$ with $\hat{\mu}_{A+} = \hat{\mu}_A + \hat{\gamma}\mu_A$
$$\hat{\gamma} = \min_{j \in A^c}^{+}\left\{\frac{\hat{C} - \hat{c}_j}{A_A - a_j}, \frac{\hat{C} + \hat{c}_j}{A_A + a_j}\right\}$$
7. Repeating the steps from the beginning for each variable selection so that all predictor variables have been selected

$\min_{j \in A^c}^{+}$ indicates that the value chosen is a positive minimum value of j which is not the set of A. In the final stage, the value $\hat{\gamma}$ using $\hat{\gamma} = \frac{\hat{C}m}{A_m}$,

B Learning Process and Model Evaluation

The method used in building in this application is Least Absolute Shrinkage and Selection Operator with Least Angle Regression is a multiple linear regression that suitable to be applied to calculation of equation model when data held have outliers and eliminating variable that held errors. This application requires input data in form of Landsat 7 and Landsat 8 imageris for the Jakarta's buffer cities areas. Landsat 7 imagery requires preprocessing in the form of Landsat gap fill and dark subtraction while Landsat 8 imagery only needs preprocessing dark subtraction. Landsats imagery that will be used on the system can be downloaded on the United States Geological Survey (USGS) Earth Explorer webpage (https://earthexplorer.usgs.gov/). The Landsat Imageries requires two intersection files with map position 122064 and 122065. Bekasi, Bogor, Depok, and Tangerang areas will be used as a case study in the making of the application program to detect the land use change. The map that is used as data in land use change detection are band blue, band green, band red, band near infrared radiation (NIR), band shortwave infrared radiation (SWIR-1 and band SWIR-2. Band blue, red, green represent a RGB images that are easily captured by human eyes. Band NIR uses to detect vegetation density when combined with band red or known as NDVI. Band SWIR-1 and SWIR-2 use to detect vegetation using atmospheric penetration when combined with band red.

This study is based on previous research from Herwindiati et al., "Impervious Surface Mapping Using Robust Depth Minimum Vector Variance Regression." [11], and "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" Landsat imageries." [12]

The model is trained using training data that has been labeled into each class. The classes are green land, partial green land, and impervious land. The program will calculate the average NDVI score for each class. The NDVI score will be used as variable estimator. NOVI values range between -1 to 1. The value -1 indicates that the area is an impervious land, while the value 1 indicates that the area is green land. NDVI value will be calculated as follows:

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$$

where:
NIR = band 5 of Landsat 8 imagery
Red = band 4 of Landsat 8 imagery

After that the model will be trained using landsat imageries on the band blue, band green, band red, band NIR, band SWIR-1 and band SWIR-2 as its parameters. After the models are formed, then an evaluation of the model is carried out by selecting the model with the smallest mean square error and seeing the best model's precision and recall score using confusion matrix. Based on the LASSO LARS algoritm, as example, imperviois estimator variable models that have been made can be seen in the following tables dan figures:

Table 1 Estimator of the LASSO coefficient for Impervious Land Class Landsat 8

Step	X1	X2	X3	X4	X5	X6
0	0	0	0	0	0	0
1	-0.2035	0	0	0	0	0
2	-1.1839	0	0	0.9803	0	0
3	-0.1303	0	-4.1748	4.1097	0	0
4	-0.1369	0	-4.2280	4.1801	0	-0.0115
5	-0.1437	0	-4.2259	4.1917	-0.0184	0

31

6	-0.1941	0	-4.2103	4.2573	-0.0509	0
7	-0.2002	0.0269	-4.2323	4.2577	-0.0499	0
8	-0.5081	1.0638	-5.0599	4.4689	-0.5826	0.4309

In the table 1, LASSO LARS algoritm estimated for every coefficent that used in the model and stop when the algortim has found the best model based on the smallest mean square error that can be seen in Figure 1 for the Landsat 7 data and Figure 2 for the Landsat 8 data.
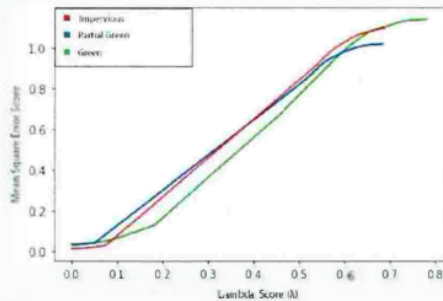

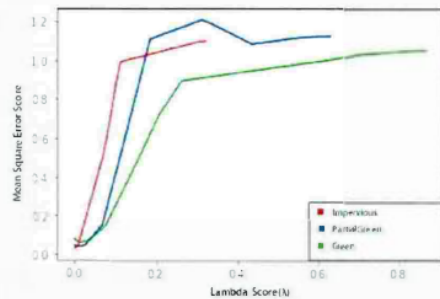Figure 1 Graph of Lambda Value with Mean Square Error Score Landsat 7

Figure 2 Graph of Lambda Value with Mean Square Error Score Landsat 8

The best models that were formed were 6 models, the models are Landsat 7 Green Model, Landsat 7 Partial Greeb Model, Landsat 7 Impervious Model, Landsat 8 Green Model, Landsat 8 Partial Green Model, and Landsat 8 Impervous Model. The best model alpha and MSE score for Landsat 8 data can be seen on Table 2. Those models will be used on testing data to calcute the accuracy and F1-Score. The classification is done by selecting the class that has the smallest difference between model score and estimator variable (NDVI) from each class. The accuracy can be calculate using confusion matrix that can be seen in figure 3

Table 2 The Best model selection of LASSO Regression on Landsat 8

No	Class	Coeficient Variables	Alpha	MSE
1	Green Land		0.70391	1.1131
			0.20172	0.1503
			0.02471	0.036
		$0X\_1 + 0X\_2 - 1.6343\,X\_3 + 1.20519\,X\_4 - 0.07644\,X\_5 + 0\,X\_6$	0.01347	0.034
2	Partial Green Land	$0X\_1 + 0X\_2 - 1.39826\,X\_3 + 1.19862X\_4 + 0X\_5 - 0.00683\,X\_6$	0.51902	0.7989
			0.31445	0.3286
			0.02173	0.0355
			0.01762	0.0349
3	Impervious Land	$-0.50811\,X\_1 + 1.06377\,X\_2 - 5.05998\,X\_3 + 4.46887\,X\_4 - 0.58264\,X\_5 + 0.43098\,X\_6$	0.31532	0.3303
			0.11179	0.0681
			0.07833	0.0507
			0.00348	0.0345
			0.00186	0.0345
			0.00176	0.0344
			0.00063	0.0337
			0.00061	0.0337
			0.	0.0334

From Table 2, the LASSO LARS algorithm removes the blue, green, and SWIR-2 band from Landsat 8 green land model and remove the blue and green band from Landsat 8 partial green land model. Those variables are eliminated because they hold error for the model.
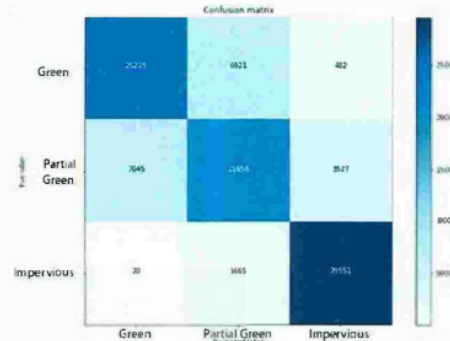

Figure 3 Confusion Matrix LASSO Regression on Landsat 7

From the confusion matrix above, the diagonal part shows the accuracy between the predictions and the actual conditions. while other areas on the same line indicate the prediction error of the actual conditions. As a result, the average accuracy of Landsat 7 models is 77.52% and average accuracy of Landsat 8 models is 81.14%. The average accuracy and F1-Score can be seen on Table 3.

Table 3 Average Accuracy and Average F1-Score on Landsat 7 and Landsat 8

	Average Accuracy(%)			Average F1-Score(%)		
	G	PG	I	G	PG	I
Landsat 7	77.31	67.2	94.6	77.71	69.33	91.21
Landsat 8	77.3	50,67	93,47	72.66	59.37	84.52

Where:
G = Green
PG = Partial Green
I = Impervious

III. Result and Disccusion

Testing with three classes (green, partial green, impervious) will use testing data for the Greater Jakarta area in the 2008, 2011, 2014, 2017 and 2020. In image detection using Landsat 7 and Landsat 8 imageries, the resolution of each pixel represents 30m x 30m the original area. Every one-pixel imagery will represent 0.0009 km². The result of calculation land area based on each district in 2008, 2011, 2014, 2017 and 2020 with observation of 3 classes can be seen in Figure 4.

Figure 4 Calculation of Land Area in 3 Classes in 2020

Based on the calculation of land use change, ranking of land changes from green and partial green to impervious will be carried out to observe the biggest land changes in subdistrict areas. An example of land use changes ranking that occurred in Depok City from 2014 to 2017 can be seen on Figure 5.

Figure 5. Land Use Change Ranking in Depok City from 2014 to 2017

Based on Figure 5, Cinere district has the highest land use change that happened between 2014 to 2017. The land use changes ranking that occurred between Regency/City can be seen on Figure 6
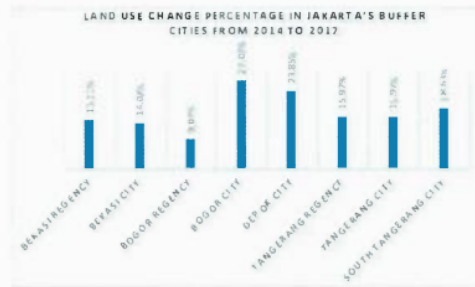

Figure 6 Land Use Change Percentage in Jakarta's Buffer Cities from 2014 to 2017

Based on Figure 6, Bogor city has the highest land-use change that happened between 2014 to 2017 which is 27.03% of its area. Based on the calculation of land area, an analysis of land changes is carried out from year to year. An example of the land use change mapping in the Cilodong sub-district from 2008 to 2020 can be seen in Figure 7 and Table 4 and the land use change to impervious land mapping in the Depok City can be seen on Figure 8.
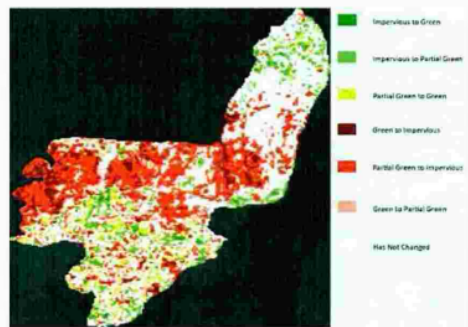

Figure 7. Land Use Change Mapping in Cilodong Subdistrict, Depok Ciry from 2008 to 2020

Table 4. Calculation of Land Use Change in Cilodong Subdistrict, Depok City

First year	Last Year	Land Use Change Area (km²)						Land Use Change Percentage(%)				
		G to PG	G to I	PG to G	PG to I	I to G	I to PG	G to PG	G to I	PG to G	PG to I	I to G
2008	2011	1.0755	0.1944	0.2745	2.0601	0.0099	0.8910	6.48	1.17	1.65	12.41	0.06
2008	2014	0.6993	0.2628	1.2744	1.3212	0.1440	2.0646	4.21	1.58	7.67	7.96	0.87
2008	2017	0.7938	0.2691	1.0782	1.9782	0.1584	1.4562	4.78	1.62	6.49	11.91	0.95
2008	2020	0.4617	0.9324	0.8280	4.7682	0.1719	1.0962	2.78	5.62	4.99	28.72	1.04
2011	2014	0.2151	0.0774	1.3455	1.2528	0.1062	1.9449	1.30	0.47	8.10	7.54	0.64
2011	2017	0.1305	0.5112	1.0494	4.2516	0.1836	1.5003	0.79	3.08	6.32	25.60	1.11
2011	2020	0.1701	0.0927	1.6029	0.8883	0.1017	2.8575	1.02	0.56	9.65	5.35	0.61
2014	2017	0.5193	1.1178	0.6642	4.7025	0.1224	0.6408	3.13	6.73	4.00	28.32	0.74
2014	2020	0.7884	0.0774	0.5076	1.8351	0.0756	0.5796	4.75	0.47	3.06	11.05	0.46
2017	2020	0.3375	0.9882	0.7308	3.7503	0.0270	0.9117	2.03	5.95	4.40	22.59	0.16

Where:

G = Green
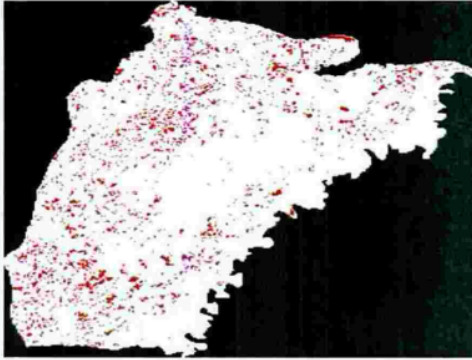
PG = Partial Green

I = Impervious

Figure 8. The Land Use Changes to Impervious Land of Tapos Subdistrict in Depok City from 2014 to 2017

Based on the calculation of land classification in 2020, it can be colcluded that the amount of green spaces area on each regency/city can be seen in Figure 9
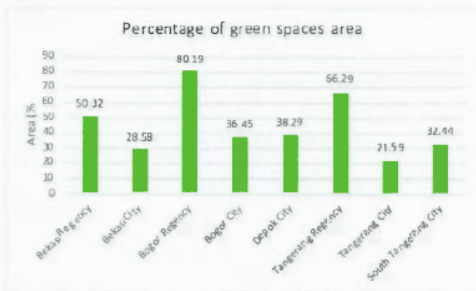

Figure 9 Percentage of Land Use Change each Regency/City in 2020

As a summary, the calculation of land area every 3 years that happened in each Regency/City can be seen on Table 5 and the mapping of land classification in the Jakarta's buffer city can be seen on Figure 10

Figure 10 The mapping of land classification in the Jakarta's buffer city in 2017

Table 5. Calculation of Land Area for 3 Classes Each Regency/City From 2008 to 2020

No	District	Year	Green (km²)	Partial Green (km²)	Impervious (km²)
1	Bekasi Regency	2008	72.9486	304.0056	894.2643
		2011	317.4849	394.1622	559.5714
		2014	237.5523	510.2883	523.3761
		2017	158.8329	516.4848	595.8990
		2020	228.4128	411.3162	631.4877
2	Bekasi City	2008	10.4175	63.1260	142.8453
		2011	11.8890	65.0241	139.4757
		2014	13.0518	48.7944	154.5426
		2017	1.9620	51.2640	163.1628
		2020	7.0353	64.4625	144.8910
3	Bogor Regency	2008	1648.4157	954.5643	394.9227
		2011	1827.4338	895.3299	275.1399
		2014	1723.8456	926.2242	347.8347
		2017	1674.4968	1011.3390	312.0696
		2020	1387.4220	1016.5077	593.9748
4	Bogor City	2008	16.1793	52.1874	45.2349
		2011	38.7054	55.7055	19.1907
		2014	23.2731	47.3499	42.9786
		2017	23.7015	45.7488	44.1513
		2020	13.4541	27.9513	72.1962
5	Depok City	2008	14.7825	98.7471	88.2639
		2011	11.0673	92.0070	98.7192
		2014	32.5215	95.3658	73.9053
		2017	29.8089	83.1204	88.8633
		2020	21.9573	55.3140	124.5213
6	Tangerang Regency	2008	109.5462	447.7716	472.7979
		2011	78.3585	428.3514	523.4058
		2014	235.5867	488.4588	306.0720
		2017	167.0103	473.4774	389.6298
		2020	314.9163	367.9965	347.2047
7	Tangerang City	2008	9.2718	30.1473	143.1864
		2011	5.5395	45.3510	131.7096
		2014	4.4280	32.8536	145.3185
		2017	9.1863	47.3553	126.0639
		2020	8.6049	44.5815	129.4191
8	South Tangerang City	2008	3.0645	53.7345	108.7236
		2011	14.5530	69.9066	81.0621
		2014	12.5748	58.6521	94.2948
		2017	14.0499	39.6459	111.8259
		2020	6.0597	69.9435	89.5185

IV. Conclusion

Based on the testing data, the LASSO LARS model can classify land into three classes with average accuracy 81.14% on Landsat 7 imagery and 77.52% on Landsat 8 imagery with F1-Score for each class are 93.63% on green land, 67.49% on partial green land, and 81.99% on impervious land. Based on data classification, the biggest impervious land is occurred in Tangerang City that has 78.41% area that classified as impervious land with the biggest land use changes happened in Neglasari subdistrict 17.51%. Percentage of the amount of impervious land and non-impervious land every 3 years experiences fluctuating changes. Within the Jakarta's buffer cities, Bekasi Regency, Bogor City, Bogor Regency, Depok City, and Tangerang City have the numbers of impervious land has increase in 2020 while Bekasi City, Tangerang Regency, and South Tangerang City have the number of impervious land increase in 2017. The following are the land mapping in Tangerang City in 2020 and its comparison with ground truth imageries with Google Earth can be seen in Figure 11.
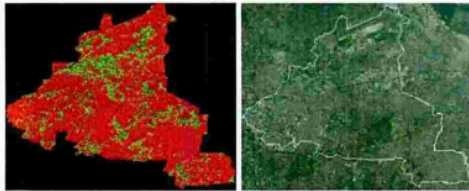
Figure 11. Comparison of Land Mapping and Ground Truth Imagery on Tangerang City in 2020

According to Table 5, Bekasi City and Tangerang City need to increase their green space areas by looking at the subdistrict that have most impervious land, so it will increase the green space area so they can achieve the minimum standart that written on Law Number 26 Year 2007.

REFERENCES

[1] R. Scalenghe and F. Ajmone-Marsan, "The anthropogenic sealing of soils in urban areas," *Landsc. Urban Plan.*, vol. 90, no. 1–2, pp. 1–10, 2009, doi: 10.1016/j.landurbplan.2008.10.011.

[2] Badan Standardisasi Nasional, "Klasifikasi Penutup Lahan," *Sni 7654*, pp. 1–28, 2010.

[3] N. Geographic, "Urbanization." https://www.nationalgeographic.org/encyclopedia/urbanization/ (accessed Jan. 21, 2021).

[4] Badan Pusat Statistik, "Penduduk Indonesia Hasil Survei Penduduk Antar Sensus 2005," *Badan Stat. Indones.*, p. 634, 2005.

[5] Badan Pusat Statistik, "Penduduk Indonesia: Hasil Sensus Penduduk 2010," *Badan Stat. Indones.*, p. 706, 2010. [Online]. Available: https://www.bps.go.id/.

[6] Badan Pusat Statistik, "Penduduk Indonesia Hasil Survei Penduduk Antar Sensus 2015," *Badan Stat. Indones.*, p. 462, 2015.

[7] M. Z. Wahyunto, a. P. Abidin, and Sunaryanto, "Studi Perubahan Penggunaan Lahan DAS Citarik, Jawa Barat dan DAS Garang, Jawa Timur," *Semin. Nas. Multifungsi Lahan Sawah*, pp. 39–40, 2001.

[8] A. Welianto, "Kepadatan Populasi Manusia: Dampak dan Pengaruhnya," 2020. https://www.kompas.com/skola/read/2020/02/04/160000369/kepadatan-populasi-dampak-dan-pengaruhnya?page=all.

[9] R. Indonesia, "Undang-Undang Republik Indonesia No.26 Tahun 2007 Tentang Penataan Ruang," vol. 136, no. 1, pp. 23–42, 2007, [Online]. Available: https://books.google.co.id/books?id=rxKrxYYJpFkC&pg=PA108&dq=kenyamanan+ruang+terbuka+hijau&hl=id&sa=X&ved=2ahUKEwis6-773tHsAhVEAXIKHUV9CAQQ6AEwAXoECAIQAg#v=onepage&q=kenyamanan ruang terbuka hijau&f=false.

[10] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *J. R. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.

[11] Herwindiati, D.E.; Hendryli, Janson; Hiryanto, Lely,. "Impervious Surface Mapping Using Robust Depth Minimum Vector Variance Regression," *Eur. J. Sustain. Dev.*, vol. 6, no. 3, p. 10, 2017, doi: 10.14207/ejsd.2017.v6n3p29.

[12] J. Handoko, D. E. Herwindiati, and J. Hendryli, "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," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 581, no. 1, 2020, doi: 10.1088/1755-1315/581/1/012045.

DYAH E H 'LAND USE CHANGE USING LEAST' PROSIDING_0001

10% SIMILARITY INDEX

7% INTERNET SOURCES

6% PUBLICATIONS

5% STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

3%
★ www.researchgate.net
Internet Source

Exclude quotes	On	Exclude matches	< 1 words
Exclude bibliography	On		