An XGBoost Based Approach for Urban Land Use and Land Cover Change Modelling

Md Didarul Islam[1], Kazi Saiful Islam[2], and Mohammad Mia[2]

[1]Central Michigan University
[2]Khulna University

May 10, 2021

Abstract

Land use and land cover (LULC) change have significant consequences on habitat and environment. Scholars have developed several LULC models to identify the factors behind the changes and to simulate future LULC scenarios to assist in policymaking. Nevertheless, the accuracy of the models remains contentious and a matter of ongoing research agenda. Additionally, most of these studies used a training dataset to train the model and a validation dataset, which is a part of the original training dataset used to validate the model's accuracy. However, to justify model's actual predictive capability, we need to test the model on real-world dataset that was not used in modeling. So, we present XGBoost model to improve the accuracy of LULC prediction. Contrary to the typical studies, we use a separate test dataset to justify the model's predictive capacity in real-world scenario. The result reveals that XGBoost model exhibits highest 84% kappa and 93% accuracy score compared to two benchmark model LR-CA (82% kappa and 92% accuracy score) and ANN-CA (82% kappa and 92% accuracy score). We also found that the built-up area increased by 48.7% in 2002 to 64% in 2010, while agricultural and vacant land declined by almost at the same magnitude over the period and the most important aspect of the LULC shift process in Khulna city was the proximity factors to major roads, industry and commercial establishments. The proposed model proved to increase the predictive accuracy making it much more reliable for analyzing and predicting urban LULC using spatial factors.

Table 1 Selected Factors for LULC Change Modeling

No	Factors	Factors Class
1	Distance to major roads	Proximity causes
2	Distance to the central business district (CBD)	Proximity causes
3	Distance to commercial establishment	Proximity causes
4	Distance to industry	Proximity causes
5	Distance to education establishment and	Proximity causes
6	Population per 30m grid	Socio-economic factor

Parameters	Values
learning rate	0.2
Max_depth	9
Min_child_weight	10
Gamma	0.5

Table 2 Tuned Values of Selected Parameter of XGBoost model

Class	Predicted Yes	Predicted No
Actual Yes	TP	FN
Actual No	FP	TN

Table 3 Confusion Matrix Table

Here, TP= True Positive (change in both real & predicted), TN= True Negative (no change in both real & predicted), FP= False Positive (change in predicted but not in real) and FN= False Negative (change in real but not predicted).

Table 4 Area of LULC classes of year-2002, 2010 and 2018

Class	2002 Area	2002 Area	2010 Area	2010 Area	2018 Area	2018 Area	[?] (2002-2010)	[?] (2
	Sq.km	%	Sq.km	%	Sq.km	%	Sq.km	%
Agriculture & Vacant	19.63	43.3	13.49	29.6	10.37	22.9	-6.24	-13.8
Wetland	3.60	8.0	2.90	6.4	2.45	5.4	-0.70	-1.6
Built-up	22.06	48.7	29.01	64	32.48	71.7	6.94	15.3

Variables	Coef.	Std. Err.	Z	P>|z|	[0.025	0.975]	Odds Ratio	VIF
Dist. to major road	-4.4477	0.2195	-20.2632	0.0000	-4.8779	-4.0175	0.012	2.0
Population per 30 m grid	0.1122	0.0014	83.0208	0.0000	0.1096	0.1149	1.119	1.8
Dist. to major education institute	-0.6650	0.0200	-33.2897	0.0000	-0.7041	-0.6258	0.514	5.5
Dist. to Industry	-0.7099	0.0290	-24.4398	0.0000	-0.7668	-0.6529	0.492	3.7
Dist. to Commercial Centers	-2.2562	0.0624	-36.1525	0.0000	-2.3785	-2.1339	0.105	4.5
Dist. to CBD	0.3216	0.0112	28.6166	0.0000	0.2995	0.3436	1.380	6.1

Table 5 Logistic regression model result

Table 6 Kappa Score and Accuracy Score for Predicted and Actual LULC of year-2018

Model	Kappa Score (%)	Accuracy Score (%)
XGBoost	84%	93%
LR-CA	82%	92%
ANN-CA	82%	92%

Table 7 Comparison of area between observed LULC of year-2018 and Predicted LULC of year-2018

Class	Year-2018 Area (Actual)	Year-2018 Area (Actual)	Year-2018 Area (Predicted)	Year-2018 Are
	Sq.km	%	Sq.km	%
Agriculture & Vacant	10.37	22.89	12.67	27.98
Wetland	2.45	5.4	2.84	6.3
Built-up	32.48	71.7	29.77	65.7

Hosted file

`Image.pdf` available at https://authorea.com/users/348289/articles/473701-an-xgboost-based-approach-for-urban-land-use-and-land-cover-change-modelling