Gauging Migration in Bangladesh

Tiana Le, Sheeba Moghal, Ishaan Babbar, Liz Kovalchuk
Georgetown University Data Science and Analytics

Abstract

This project examines the retrospective migration data collected in southwestern Bangladesh provided by the Authors Amanda R. Carrico and Katharine Donato from their paper Extreme Weather and Migration: Evidence from Bangladesh. Their study investigates the relationship between extreme weather conditions in the region and patterns of migration. Our team's analysis of the data delves deeper into the initial findings provided by Carrico and Donato, investigating beyond the primary author's conclusion associating weather with increased migration (particularly after dry spells, albeit also after warm spells and above average rainfall but to a lesser extent). Similarly, our team's feature selection showed that internal migration could be predicted by socio-economic factors such as occupation, wages in taka, and level of education. The most effective approaches for the two analyzed datasets were ensemble models; model comparison metrics and a description of the feature tree as well as variable importance is included.

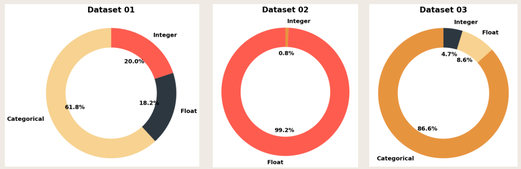
Introduction

In Bangladesh, the intertwined challenges of climate change and migration have become their own term: climate displacement. As climate change intensifies, rising sea levels and erratic weather patterns wreak havoc on the livelihoods of millions, particularly affecting those dependent on agriculture. Migratory workers, often from rural areas, find themselves compelled to leave their homes in search of alternative means of sustenance as their traditional occupations become untenable. These migrations, spurred by environmental degradation, exacerbate existing socio-economic strains, leading to overcrowded urban centers and strained resources.

Bangladesh, located in South Asia, is one of the most densely populated countries in the world. The nation is grappling with ongoing efforts to mitigate the impacts of climate change and provide support to its vulnerable communities; those impacted by the migration crisis. To address some of these challenges, data collection and analysis can help to better understand what these migrants are up against and the depth of their migration, ultimately providing opportunities to propose further research and areas for assistance.

Methodology & Data

Data was pulled from the extensive surveys commissioned by Carrico and Donato. Each of the three cleaned sets are below, the donut charts are meant to show how much of the information provided was categorical, driving our methodology and choice of instruments.

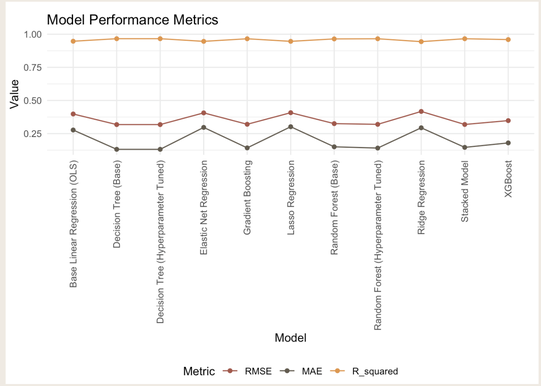


Each dataset held survey question responses, often categorically or binary encodings. The integer and float values are often of identifying information, not of numeric values. Each set was used to investigate climate change's impact on Bangladesh migratory workers.

Analysis

Dataset 01

In the analysis of the the first dataset for migration in Bangladesh, we perform feature selection using both numerical and categorical features. We aim to identify the most relevant values that can be subsetted based on our Y target variable, which is to determine if environmental change due to temperature increase/decrease is a feature of migration. To achieve this, we use ANOVA based on the f-test for numerical features, and Chi-squared test based on the p-value for categorical features. We determine the categorical feature to be significant if the p-value is less than 0.05.



The Decision Trees, Gradient Boosting, Random Forests, and Stacked Model all performed well. The R squared values are all above 0.90, indicating a strong positive relationship. RMSE for these values were approximately 0.3 or below. Their MAE are also closer to zero, showing a lower absolute error for the model's predictions after training.

Based on the analysis of the models used in the first dataset, the Decision Tree (Base) Model ended up being the best model. It contains the lowest number of RMSE (0.3181162) and MAE (0.131386) as well as the highest R-squared (0.9661455), which shows a better fit to the data compared to other models. Therefore, the Decision Tree (Base) is the best choice for predictive modeling for the first dataset. For future prediction of migration of workers as a relationship to temperature, using a decision tree or similar model should be effective.

Dataset 02

For the classification model, the survey data was subsetted to include the heads of households with the primary purpose of migration for work or to earn money.

	Specificity	Accuracy	Sensitivities	Precision	AUC
Logistic Regression	0.58	0.88	0.94	0.92	0.82
Lasso	0.67	0.89	0.92	0.96	0.89
Elastic	0.67	0.89	0.92	0.96	0.89
Decision Tree (w/ tuning)	0.48	0.90	0.99	0.91	0.89
Random Forest	0.61	0.92	0.97	0.93	0.94
XG	0.78	0.94	0.98	0.95	0.98

Analysis

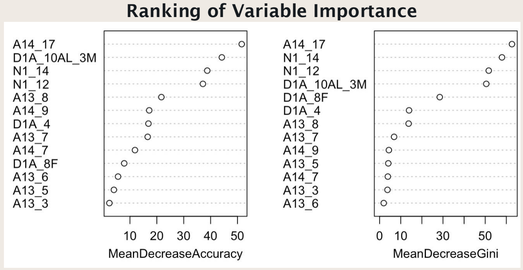
After performing logistic regression, 13 statistically significant features were identified as the predictors for the logistic regression model for classification.

Modeling

The logistic regression base model showed that the total number of trips of the head of household, rent per month, and average monthly remittances sent home were predictors of the target question (did survey respondents migrate for work or to earn money). There was also a relationship to "wage" in taka. Duration of stay, level of education (at five different levels), respondents Livelihood/occupation, were also were predictors of migration.

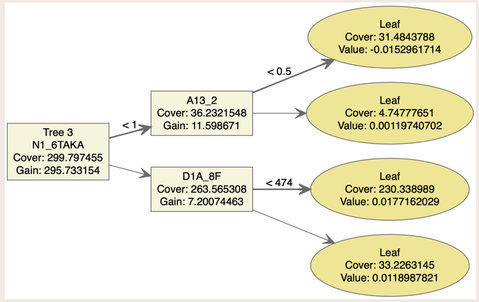
These features were selected for the final logistic regression model and repeated in lasso regression, elastic regression, decision tree with hyper parameter tuning, random forest, and XGBoost. These features predicted the primary purpose of the trip was to work or earn money. These predictors include total number of trips, rent per month, average monthly remittances, monthly wage, levels of education, and occupations, including rickshaw driver, construction worker, non-agricultural workers (factory worker, blue collar service), and homemaker.

Ranking of Variable Importance



A14_17, occupation is homemaker, had the highest variable importance. D1A_10AL_3M is monthly wage. N1_14 and N1_12 are average monthly remittances and rent per month. The A13 category is household education level.

XGBoost Tree



Dataset 03

The third survey's unique questionnaire data contained health information about respondents diagnosis, their visits to physicians and other forms of seeking medical care, or recurring symptoms. Many respondents did not report such data, and gauging the status of "health" before and after migration experiences was ephemeral to categorize within the survey. As a result, much of this data was discarded to focus on the primary modeling from the first and second datasets.

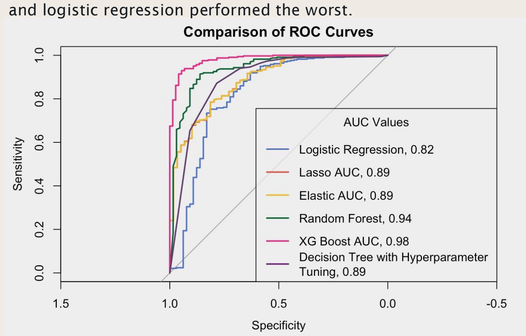
Results & Conclusion

Dataset 01

The first dataset showed an impact of house services, local government effectiveness, and environmental changes profoundly influences migration in Bangladesh. Poor infrastructure exacerbates the hardships of climate change, prompting migration when basic needs are unmet. Altered rainfall patterns and increased flood severity directly affect agriculture and habitation, driving people to relocate in search of stability. The abundance of fish, vital for many livelihoods, also shifts with changing climates, further influencing migration decisions. These factors collectively highlight how climate change intricacies crucially shape migration dynamics in regions like Bangladesh. Effective local governance can mitigate such impacts, potentially reducing displacement.

Dataset 02

The performance of the models can be observed with the ROC curve and AUC. From both the AUC and ROC, XGBoost performed the best and logistic regression performed the worst.



Given the primary factors for importance from the second dataset, rent and remittances draw attention to domestic economic challenges, which spur migration as a means of easing financial circumstances. Earning potential, a key factor driving migration, is strongly correlated with wage levels. Job chances are influenced by education, and certain occupations—such as construction or rickshaw driving—often correlate with employment opportunities in cities, which encourages migration. All of these factors contribute to the decision-making process regarding relocation for economic improvement.

Further Research

Our analysis showed certain patterns in Bangladesh for migration; further research would be to follow migration patterns in similar demographic or neighboring countries. As climate change and other economic challenges occur as a result of rising temperatures, following migratory patterns of individuals and understanding why they choose to relocate or are relocated is likely to become more and more relevant. The team's analytical approach can likely be replicated for further research in migration.

Acknowledgements

- PresentationGo. (2024). Urban Sunset: Custom Color Palette for PowerPoint. PresentationGo. https://www.presentationgo.com/presentation/urban-sunset-custom-color-palette-for-powerpoint/
- A.R., Donato, K. Extreme weather and migration: evidence from Bangladesh. Popul Environ 41, 1–31 (2019). https://doi.org/10.1007/s11111-019-00322-9