Do OOSY What I See?:
An Analysis of the Out-of-School Youth Rate in the BARMM

Arnold Zamora, Paola Basinillo

[1] arnoldjanssen.zamora@teradata.com, [2] paola.basinillo@teradata.com

Abstract—**The BARMM is a region that struggles when it comes to education, having the highest OOSY rate among Philippine regions. This study aims to examine OOSY rate for ages 5-9, 10-14, and 15-19 in the region, and the factors affecting them. OOSY rate was related to education, demographic, and socioeconomic features. Ordinary least squares (OLS), spatial regression, and XGBoost models were fit and compared. Three prominent features appeared in all final models: floor area per household, no. of teachers per school, and proportion of married male youth. The selected features per age group model suggest that housing and living conditions are more predictive for younger children, "adult" matters like marriage and household size are more predictive for older children. The municipalities of Marogong, Tapul, and Sitangkai were identified as outlying across all models; their OOSY rates were consistently underestimated at all age groups. Overall, results suggest that the region take a more holistic approach in the improvement of OOSY, taking into account the home environment and living conditions.**

I. INTRODUCTION

Following the ratification of the Bangsamoro Organic Law in 2018, the Bangsamoro Autonomous Region in Muslim Minsanao (BARMM), formerly the Autonomous Region in Muslim Mindanao (ARMM) faces a period of transition. The Open Bangsamoro Data Challenge is an initiative by Data Ethics PH that invites people to make use of open data to aid in decision-making for this transition. By processing the data into a simple language and format, it can help both decision-makers and citizens become more informed of their situation [1].
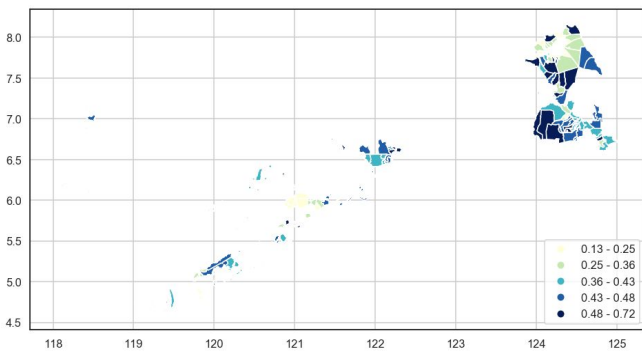

Fig. 1. OOSY Rate of Bangsamoro Autonomous Region in Muslim Mindanao for Ages 15-19 (2015)

With a complex history of conflict and displacement, the BARMM has struggled to protect basic education. Out-of-School Youth (OOSY) are children of school-going age (5-19 years old) that are currently not attending school, based on the Philippine Statistics Authority Population Census, as seen in Fig. 1. The region is consistently the lowest performing in the country when it comes to education access and retainment [2], with an OOSY rate of 14.4% against a nation average of 9.1% [3].

Studies have identified socioeconomic, demographic, and physical access factors as determinants of education participation [4, 5, 6, 7]. Specifically, the most common reasons for OOSY in ARMM are related to marriage or family matters and poverty [2]. As for the behavior of such data, education, socioeconomic and conflict factors have been shown to display spatial dependency [8, 9].

This study generally aims to model the OOSY rates of three age groups (5-9, 10-14, and 15-19) against various education, demographic, and socioeconomic features. Specifically, the study aims:

1) to perform an exploratory and descriptive analysis of OOSY rate,
2) to test for and examine nature of spatial autocorrelation of OOSY rate,
3) to fit models to determine what factors affect OOSY rate the most and how, and
4) to identify outlying municipalities where the models underperform.

II. METHODOLOGY

A. Data & Scope

The OOSY rate was sourced from the Annual Poverty Indicators Survey 2015. Different features were sourced from Open Bangsamoro, Department of Education, and Bangsamoro Conflict Monitoring System. Only the data prior to and during 2015 were used for the models with a per-municipality granularity. The OOSY rate was segmented into three age groups: 5-9, 10-14, and 15-19.

The feature set was classified into four categories: education, demographic, socio-economic, and others which totalled to 37 independent numeric features.

B. Feature Set

Multiple feature selection methods were used to generate different feature sets for the age groups, ultimately the output of Lasso Lars (5-9) and ElasticNet (10-14,15-19) were used. OOSY rate for younger children was more strongly correlated with basic living conditions, (housing, electricity, etc.) while the rate for older children was more strongly correlated with adult matters (marriage, youth labor, etc.).

C. Models

Three models were used to model the dependent: a) ordinary least squares - simple linear regression model without spatial effects, b) general method of moments lag -

linear regression with the effects of neighboring values, c) gradient boosting (XGBoost) - non-linear model without spatial effects.

D. Results

Given below are the actual and the predicted OOSY of the models, particularly in Lanao Del Sur and Maguindanao.
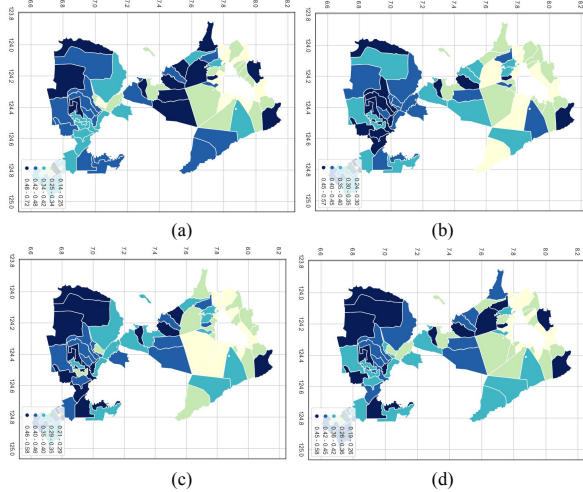

Fig. 2. OOSY Rates for 15-19. a) Actual b) OLS c) Spatial d) XGBoost

As seen in Fig. 2., the output of the XGBoost is closest to the actual OOSY, followed by the Spatial and OLS method. The adjusted R^2 and Root Mean Squared Error (RMSE) were used in evaluating the performance of the models, which can be seen in Fig. 3.
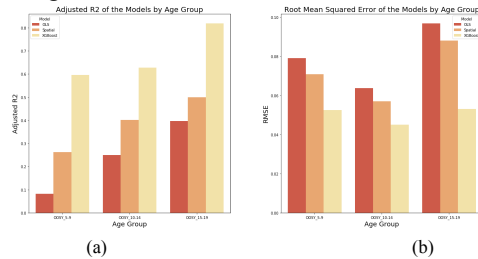

Fig. 3. Model Performance Comparison. a) Adjusted R^2 b) RMSE

After modelling, three outlying municipalities were identified using the modified z-score *(threshold: 1.5)* of the model residuals. Marogong in Lanao del Sur, Tapul in Sulu, and Sitangkai in Tawi-Tawi stood out as those whose OOSY rate values were consistently underestimated across all models and age groups. That is, these municipalities are consistently doing considerably worse than expected by the models, possibly due to outlying values of features or factors unaccounted for by the models.

III. Conclusion

OOSY Rate exhibits only weak to moderate clustering in space. This rate is dependent on independent features as well as neighboring municipality OOSY Rates and independent features. However, XGBoost outperformed the spatial model, and thus the non-linearity of the XGBoost model may be indicative of non-linear relationships in the data that are more predictive than spatial behavior.

Generally, the strongest features related to housing conditions, youth marriage, and school quantity or quality. The three most prominent such features being Floor Area, Teachers per School, and Male Youth Marriage Proportion. While school quality does appear to affect OOSY Rate, it is only one part of the story.

Comparing the age groups and feature sets, it may be vital to focus on improving the OOSY Rate for those aged 5-9 via increasing no. of elementary schools, since this has been shown to still affect OOSY for ages 10-14 and 15-19.

IV. Recommendations

To improve the models, features that weren't covered may be included (e.g. physical accessibility of schools, environmental factors such as terrain, etc). More recent data would improve how timely and relevant the models are. For the modelling itself, more complex, non-linear models that also account for spatial effects may be explored, as this may reflect the nature of the data better.

It may be useful to investigate and potentially target areas where OOSY rate is underestimated - those consistently underperforming than expected by the models. Building more schools (or making schools more accessible), increasing the number of teachers, and generally addressing poverty and living conditions may be key in improving the OOSY rate. This is consistent with a previous study in the region that suggests addressing poverty and vulnerability at the root is key in addressing OOSY [7]. Overall, results suggest that a more holistic approach be taken in development, taking into account a child's home environment asides from their school environment.

References

[1] Open Bangsamoro: Open Data for the Transition from the ARMM into the BARMM. Retrieved from: https://www.openbangsamoro.com/
[2] UNICEF. (2018). Situation Analysis of Children in the Philippines. Retrieved from: https://www.unicef.org/philippines/media/556/file
[3] Philippine Statistics Authority. (2019). Out-of-School Children and Youth In the Philippines: Educating Out-of-School Prevalence Among Filipino Children and Youth. Retrieved from: http://openstat.psa.gov.ph/Portals/0/downloads/OOSCY.pdf?ver=2018-11-10-001846-003&timestamp=1553672015112
[4] Jordan, J.L., Kostandini, G., & Mykerezi, E. (2012). Rural and Urban High School Dropout Rates: Are They Different?. Journal of Research in Rural Education. 27(12).
[5] Lyttle-Burns, A. (2011). Factors That Contribute to Student Graduation and Dropout Rates: An In-Depth Study of a Rural Appalachian School District.
[6] Tabuchi, T., Fujihara, S., Shinozaki, T., & Fukuhara, H. (2018). Determinants of High-School Dropout: A Longitudinal Study in a Deprived Area of Japan. Journal of Epidemiology. 28(11). 458-464.
[7] Philippine Institute of Development Studies. (2018). Out-of-School Children: Changing Landscape of School Attendance and Barriers to Completion. Retrieved from: https://pidswebs.pids.gov.ph/CDN/PUBLICATIONS/pidsdps1825.pdf
[8] Karahasan, B.C. & Uyar, E. (2009). Spatial Distribution of Education and Regional Inequalities in Turkey.
[9] O'Loughlin, J. & Raleigh., C (2007). Spatial analysis of civil war violence. Handbook of Political Geography.

Authors' Profile

Arnold Zamora is currently an Analytics Consultant from Teradata GDC Philippines. He finished his undergraduate degree in Computer Engineering in 2019 from the University of the Philippines - Diliman.

Paola Basinillo is currently working as an Analytics Consultant for Teradata GDC Philippines. She received her degree in BS Statistics from the University of the Philippines - Diliman in 2019, where she co-authored a paper on spatial statistics, a particular field of interest to her, in relation to voting patterns in the Philippines.