Research

Random Forest Algorithm for the Identification of Factors Contributing to Food Insecurity

Jaehyun Ahn

979-458-0193

hyun81219@tamu.edu

Department of Agricultural Leadership, Education & Communications

Texas A&M University

600 John Kimbrough Blvd

2116 TAMU

College Station, TX 77843

Alexis Zickafoose

937-509-1768

alexis.zickafoose@ag.tamu.edu

Department of Agricultural Leadership, Education & Communications

Texas A&M University

600 John Kimbrough Blvd

2116 TAMU

College Station, TX 77843

Mathew Baker

979-845-6944

Mathew.Baker@ag.tamu.edu

Department of Agricultural Leadership, Education & Communications

Texas A&M University

600 John Kimbrough Blvd

2116 TAMU

College Station, TX 77843

Random Forest Algorithm for the Identification of Factors Contributing to Food Insecurity

Introduction

Despite 33 million smallholder farms providing 70% of the food in Africa, they are among the most impoverished and food insecure (Doss et al., 2018; IFAD, 2017). In terms of gross national income Ghana ranks as an "average" African country (World Bank, n.d.). However, even in an average country, food insecurity exists as a severe problem and thus aligns with the AAAE National Research Agenda Priority *Addressing Complex Problems* (Roberts et al., 2016). Norman Borlaug (2008) emphasized the need of plant pathology, science, and technology to merge agricultural production with agribusiness strategies and agricultural education and extension to combat the divide of food insecurity. By uniting with producers and consumers, agricultural extension professionals can work to systematically improve food security (Beckman & Smith, 2008; Brown et al., 2015; Mukembo et al., 2015; Warner, 2017). Identifying predictors of food insecurity and their importance can better enable extension professionals to focus on the most dire needs.

Conceptual Framework

The conceptual framework for this study was the Needs Assessment Cycle posited by Graves (2000). This model contains six steps including deciding what information to gather, deciding when, from whom, and how to gather information, gather information, interpret information, act on information, and evaluate the effects of the action. This study involves the first four steps of Graves's (2000) model but focuses on the interpretation of the information using a novel machine learning analysis approach to interpret the information.

Methodology

The binary target categories (i.e., food insecure and food secure) in Ghana were identified by an in-country survey conducted September 2012 to March 2013 by Texas A&M University that classified food insecure households. Four districts in Ghana were purposively selected based on different agro-climatic, ecological conditions, socio-economic status, and farming practices. A total of 644 Ghanaian households were surveyed. More than 64% of households had no food shortage in the last 12 months, and among them, only 367 satisfied food availability and access based on the survey. Those became food-secure households, and the other 277 were food-insecure. A direct relationship was hypothesized between farm costs, returns, off-farm income, and food security. Income calculations included crops, livestock, off-farm, and credit. Climate, community support, market, head of household gender, household dependents, technology, agricultural information, consumption of animals, location, presence of land conflict, and farm labor-force were identified as food insecurity predictors.

Machine learning was used to prioritize the factors leading to food insecurity and confirm observed field data. To this end, *random forest algorithm* was used due to the binary nature of the data for each country. Random forest uses the training data to identify variable importance and builds many single decision trees randomly, then combines all the predictor information and selected the most voted classification (Breiman & Cutler, 2003). Misclassification rate is determined by the out-of-bag (OOB) error rate computed on the training data and test error as 150 trees built in a forest of which 50% of the entire data was used for training and 30% for random testing. Gains curves present the total votes, which in conjunction with the classification matrix allows for clearer interpretations of the data.

Results

The most important predictor identified through the Random Forest algorithm was agricultural information. If households did not have access to agricultural information, they were more likely to be food insecure. The other predictors for those without agricultural information were market access and the amount of money spent on food. For households that did have access to agricultural information, those in the Northern region of Ghana were more likely to be food insecure, especially if they spent less than or equal to $99.71 on food. For those in Ejura, Ga West, and Atwima, if they had a flood and less than or equal to 9 outside laborers, they were more likely to be food insecure. Households that had more than 9 outside laborers were more likely to be food insecure if they earned less than or equal to $791.92 from their crops. Of the total 211 households in the test data, 80 were correctly classified as food insecure, and 110 as food secure. For food security the most important variables from the OOB data were crop income, money spent to maintain those crops, money spent on food, constraints, outside labor, off-farm income, and technology usage. The OOB data had an overall misclassification rate of 9% and a 0.01 standard error. The gains curve confirmed the findings of the random forest algorithm.

Conclusions

Ghana showed that despite most of the households being classified as food secure, food insecurity is still an issue within the country. Extensionists should develop options that reflect the needs of food-insecure households could increase farming capacities and reduce food insecurity. The identified factors that most contribute to food insecurity in Ghana were lack of access to agricultural information, market access, amount of money spent on food, number of outside laborers, and crop income. Ghana's random forest presented predictors indicating farming households were independent producers, but the need of outside labor for future agriculture is emphasized due to its presence in both food insecurity and security predictors. Due to the relatively low misclassification rates, random forest can guide site-specific programming to alleviate food insecurity. This study highlights the importance of adhering to the Needs Assessment Cycle and allows for the last two steps, act on information and evaluate its effects, to be completed (Graves, 2000). Agriculture extension professionals have a structure with which they can improve food security in Ghana.

Implications

Extensionists should focus on providing agricultural information, technology education, and market access to smallholder farms in Ghana. By creating site-specific programs based on the needs of the locals, education and extension programs may have a better chance of applicability and sustainability. National organizations could use the same methods to identify food security priorities in their own communities. Retroactive learning from developing countries can better inform policy and programming in the United States. Decision tree classifiers are underutilized in the context of food security calling for further research. More evidence from more cases over extended periods is needed. A timely, systematic, and just study is necessary for international and national rural communities whose production and social environment is changing rapidly.

References

Beckman, L. L., & Smith, C. (2008). An evaluation of inner-city youth garden program participants' dietary behavior and garden and nutrition knowledge. *Journal of Agricultural Education, 49*(4), 11-24.

Breiman, L., & Cutler. A. (2003). *Random forests*. Department of Statistics, University of California, Berkeley. https://stat.berkeley.edu/~breiman/forests/cc_home.htm

Borlaug, N. E. (2008). Nobel peace laureate Norman Borlaug endorses food security. *Food Security*. https://springer.com/journal/12571/updates/17238148

Brown, N. R., Roberts, R., Whiddon, A. S., Goossen, C. E., & Kacal, A. (2015). Paxton revisited: The essence of the lived experiences of urban Agricultural Education students. *Journal of Agricultural Education*, *56*(1), 58-72. https://doi.org/10.5032/jae.2008.04011

Doss, C., Meinzen-Dick, R., Quisumbing, A., & Theis, S. (2018). Women in agriculture: Four myths. *Global Food Security*, *16*, 69-74. https://doi.org/10.1016/j.gfs.2017.10.001

Food and Agriculture Organization of the United Nations (FAO). (2007). *Food Security*. FAO. http://fao.org/fileadmin/templates/faoitaly/documents/pdf/pdf_Food_Security_Cocept_Note.pdf

Graves, K. (2000). *Designing language courses: A guide for teachers*. Heinle & Heinle.

International Fund for Agricultural Development (IFAD). (2017). *The Field Report*. IFAD. https://ifad.org/thefieldreport

Mukembo, S. C., Edwards, M. C., Ramsey, J. W., & Henneberry, S. R. (2015). Intentions of young farmers club (YFC) members to pursue career preparation in agriculture: The case of Uganda. *Journal of Agricultural Education*, *56*(3), 16-34. https://doi.org/10.5032/jae.2015.03016

Roberts, T. G., Harder, A., & Brashears, M. T. (Eds). (2016*). American Association for Agricultural Education national research agenda: 2016-2020*. Gainesville, FL: Department of Agricultural Education and Communication

Warner, L. A., Harder, A. M., Henry, C. V., Ganpat, W. G., & Martin, E. (2017). Factors that influence engagement in home food production: Perceptions of citizens of Trinidad. *Journal of Agricultural Education*, *58*(3), 239-255. https://doi.org/10.5032/jae.2017.03239

World Bank. (n.d.). *Gross national income per capita*. https://data.worldbank.org/indicator/NY.GNP.PCAP.PP.CD?most_recent_value_desc=false