Simplifying food security: Comparing predictive models based on the Global Food Security Index

Nathaniel Cinnamon David Ngo Amy Ozee Charlie Talyor

June 10, 2021

Abstract

Food security is a global issue that affects many developing countries. Households vulnerable to seasonal food security are expected to be impacted by future crises such as climate change, economic recessions, and pandemics. Past literature apply parametric models to measure food security; machine learning algorithms can improve the modeling accuracy of food security predictions. In this paper, we compare ordinary least squares, ridge regression, and random forest for predicting food security based on the Global Food Security Index (GFSI). We evaluate the model with the best out of sample accuracy and determine key indicators of the GFSI. We find a random forest provides the best out of sample predictive accuracy, with the second lag (2 years prior) of average dietary energy supply adequacy found to be the main driver of the random forest model. This paper illuminates the advantage of machine learning algorithms in predicting food security and motivates informed policy considerations for long term intervention.

Keywords: Food Security, Global Food Security Index, Latin America and Caribbean Region, Ordinary Least Squares, Predictive Modeling, Random Forest Regression, Ridge Regression

1 Introduction

As of 2018, 42 million individuals live in a food insecure situation in Latin American and Caribbean (LAC) countries (Salazar and Muñoz, 2014). With a predicted increase of 1.7 billion in world population by 2050, global food production must increase by 70% to maintain current levels of food security (Mc Carthy et al., 2018). Food insecure households can be determined by low education, limited social capital, and living in countries with low GDP per capita (Smith et al., 2017); access to food and water is also driven by economic factors such as income, livestock, and produce; and external factors such as climate change, international trade, and urbanization (Corral et al., 2000). Obtaining empirical information on food security is critical for early famine monitoring, informing effective government policy and directly targeting the causes of hunger and starvation. (Weber et al., 1988).

Countries and global organizations have developed differing approaches towards assessing food security by regional and country levels; the Global Food Security Index (GFSI) assesses country-level trends in food security, while others such as the Famine Early Warning Systems Network (FEWS NET) projects future levels of food security (Jones et al., 2013). As many indices apply qualitative and quantitative techniques in the development of their models, research has been done to identify the most relevant determinants of food security (Smith et al., 2017). Governments often lack the mechanisms to make informed decisions; past literature have attempted to provide economic and quantitative models suitable for modeling and predicting food security (Mbukwa, 2013). The issues with such models include data availability and limitations in chosen indicators being applicable outside of their studied regions.

The purpose of this paper is to develop and train a model to predict food insecurity in several countries within the LAC region and compare the model with conventionally used regressions. We apply indicators from the World Bank and Food and Agriculture Organization to construct several models that predict the GFSI score as a benchmark for food security levels. We construct least squares regression, ridge regularization, and random forest models to determine the most accurate method for predicting GFSI score. We find that the ridge regression model predicts food security with the greatest accuracy—an out of sample accuracy of 9.20%. Our results also indicate that out of sample predictive capability improves as model complexity increases. Our results help to provide a framework for choosing a suitable quantitative model in predicting food security.

1.1 Literature Review

A multitude of socioeconomic factors impact regional food security. According to the Food and Agricultural Organization, food security is determined by four dimensions: (i) the physical availability of food; (ii) economic access to acquiring food for a nutritious diet; (iii) food utilization in which diet, water, sanitation, and healthcare satisfies nutritional needs; and (iv) stability of the other three dimensions over time (FAO, 1996). Driving indicators that are often considered include household education level, household income level, quality of diet, and living in a country with a low GDP per capita (Frongillo et al., 2017; Mbukwa, 2013; Smith et al., 2017). Other possible indicators also include birthweight, food supply, livestock, and poverty levels (Masih et al., 2017).

Regional indicators that address environmental and climate characteristics are also vital in predicting food insecurity. Climate change impacts crop productivity and food availability over time; climate variability can exacerbate food security in areas vulnerable to hunger (Wheeler and von Braun, 2013). Gbegbelegbe et al. (2014) show that extreme climates can diminish crop production on the national and international level—placing pressure on smallholder farmers who rely on a single crop for a large portion of their nutritional intake. Alpízar et al. (2020) further shows that Central American households are vulnerable to seasonal food insecurity and can be further exacerbated by climate change.

The Global Food Security Index (GFSI) is one composite measure that has monitored annual national-level food security across 113 countries since 2012 (EIU, 2019). The GFSI uses 34 indicators spanning the following domains of food security: affordability, availability, and quality and safety (Jones et al., 2013). National GFSI scores are calculated by weighting these indicators and scoring countries on a 100-point scale. Values near zero indicate extremely alarming levels of hunger; values near 100 indicate low levels of hunger (Henneberry and Carrasco, 2014). Izraelov and Silber (2019) validates the GFSI as a reasonable measure of food security; they conclude that the choice of indicators and weights selected by the EIU provide a reasonable ranking of countries by their level of food security and is close to other widely used indices.

Previous literature has focused primarily on identifying determinants of food security or evaluating indices—Mbukwa (2013) proposes a logistic regression to measure the impact of indicators on food security and Backer and Billing (2021) finds FEWS NET projections of food security in Africa to be 84% accurate. However, there is an absence in applying machine learning applications to match or surpass predictive power of indices; Razzaq et al. (2021) determined random forest to be the best machine learning algorithm in of predicting food security in rural areas of Pakistan. As machine learning methods can reduce bias and overestimation that may occur in the case of nonlinear data (Cai et al., 2018), it may be desirable to expand the applications of machine learning to predict food security across countries and regions.

Thus, we find it useful to apply machine learning algorithms to improve the evaluation of food security and address the need to make informed interventions to create sustainable solutions to food insecurity (Salazar and Muñoz, 2014; Smith et al., 2017). This paper compares several computational methods to accurately predicts food security across countries; we apply ordinary least squares, ridge, and random forest regression models based on food security indicators provided and historic GFSI data. We measure the coefficient of determination (R^2) to evaluate the out of sample accuracy of our models and feature selection to identify key indicators of food security.

Section 2 describes the GFSI and data used. Section 3 presents the regression model. Section 4 presents the empirical results for predicting levels of food security with a set of indicators. Section 5 concludes.

2 Data

2.1 Variable of Interest: Global Food Security Index

The Economist Intelligence Unit's Global Food Security Index (GFSI) assesses food security across countries (EIU, 2019). The 2019 GFSI uses 34 indicators to cover several dimensions of food security: Affordability, Availability, and Quality or Safety. To calculate the index, GFSI data is scaled to a value between zero and a hundred. The three category scores are calculated as the weighted means of the indicators; the GSFI score is then calculated as the weighted means of the category scores, where values approaching 100 represent food secure countries. Table 3 presents the GFSI correspondence to each country and year. We do not adjust these results with the optional Natural Resources and Resilience category; past literature have shown that GFSI scoring using the three default category scores provides a reasonable measure for assessing food security. In this project, we have 9 observations of GFSI Score per country for a total of 90 observations of the outcome variable (Henneberry and Carrasco, 2014; Izraelov and Silber, 2019; Thomas et al., 2017).

2.2 Explanatory Variables

We collected several food security indicators from the Food and Agriculture Organization (FAO) and World Bank. FAO data collects information on exports and imports, employment levels, per capita food supply, and potential food driven health effects on women and children for LAC countries between the years of 2010 to 2019. The World Bank provides data on industry shares of GDP, national support policies, and multidimensional poverty headcounts. The countries of interest in this project include: Argentina, Bolivia, Brazil, Chile, Ecuador, El Salvador, Guatemala, Honduras, Nicaragua, and Peru. In total there are 100 observations which equate to 10 years worth of data for each indicator in each country.

Table 1 and Table 2 display each of the indicators being used in our models, their average value and standard deviation between the years 2010 and 2019, and separate them by information source. The majority of the information gathered from these sources are results from large surveys conducted by local institutions.

2.3 Data Processing

We collected GFSI and World Bank data with the intention to have indicators of GFSI for the years 2012 to 2020. The countries we collect data for include: Argentina, Brazil, Bolivia, Chile, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Peru, Venezuela. We combine FAO and World Bank data together to observe indicator values across each country and each year. We also combine the data set with GFSI data for ease of applying regression models in later sections. For our purposes, we will use lagging data to predict the outcome (GFSI). We expect that predicting GFSI in year t requires information from years $t-1$ and $t-2$, our models will reflect this assumption and include lagging data.

While some indicators report data for each year, other indicators report data every few years. To address empty values in the data, we apply imputation—a method in which we replace missing data with substituted values. In this case, we compute substituted values by applying a countryâs year-fixed value for one indicator into subsequent years. Our method of addressing missing values in data is unique; Mbukwa (2013) analyzed food security with survey data collected in a district within Tanzania, while Razzaq et al. (2021) collected data in rural areas of the Punjab province in Pakistan. While past literature has focused on comparing rural data in a single country to study food security, we collect and impute data across countries to compare food security at the macro-scale.

3 Problem description

As discussed in the literature, classifications such as the GFSI apply dozens of variables across different dimensions to measure food security. It can be burdensome to collect the required data for all of these variables. Additionally, too many variables in a model can lead to a curse of dimensionality—requiring an extremely large amount of test data to make our model useful for out of sample prediction. We select a smaller set of relevant indicators of food security based on the availability of data and develop a model that predicts GFSI as measure for food security. We construct several different models and test the accuracy of each while using our selected set of predictors.

We first estimate an ordinary least squares regression model where the outcome variable is the GFSI score ranging from 1 to 100, and the predictors are the selected subset of indicators discussed above. GFSI is continuous and the higher scores represent greater food security. We estimate the following model:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n + \varepsilon_i,$$

where $\hat{Y}$ is the predicted or expected GFSI score, $X_1, \ldots, X_n$ are the n distinct predictor variables, β_0 is the predicted GFSI score when all predictors are equal to zero, $\beta_1, \ldots, \beta_n$ are the estimated regression coefficients, and ε_i is the error term. We assume ε_i to be normally distributed.

In the next model we apply a form of regularized regression; we apply a ridge model, a form of penalized regression that improves on OLS in making out-of-sample predictions. This technique can shrink our coefficients relative to OLS; in situations where OLS estimates have high variance, we can see a significantly reduced variance with only a small increase in bias. Ridge models are effective in cases where there are a large number of predictors compared to our sample size n, or when we expect multicollinearity in the data. In our case, we would expect that a lot of food security drivers would be correlated. Recall that for OLS, we minimize the sum of squared residuals (SSR). For the Ridge model, we minimize:

$$L_{optimal} = SSR + \alpha \sum_{j=1}^{n} {\beta_j}^2,$$

where α is a tuning parameter and β_j is the slope coefficient for indicator j. The term being added to the SSR is called the shrinkage penalty—it penalizes indicators with larger coefficients. In this model, we run cross validation with different α values and pick the one which minimizes $L_{optimal}$. In this model, we assume that each indicator has an impact on the GFSI score.

We estimate a random forest as the third model. Implementing this model is necessary because the previous two models only consider linear relationships between indicators and GFSI score; we want to consider non-linear relationships as well. A random forest is a vast array of decision trees—all with different combinations of predictors—that predicts the outcome based on the results from each tree. The individual trees are largely uncorrelated, so the outcome most frequently predicted by the decision trees becomes the model outcome. The large number of uncorrelated models acting as a committee will outperform each individual tree in the forest. A random forest model also implicitly performs feature selection: as the individual decision trees are constructed, the model keeps track of which indicators do the best job at consistently predicting the right outcome. The final model output will therefore include a ranked list of the most important predictors in determining GFSI score.

Finally, to evaluate each of the three models evenly, we will perform a train-test split on the data to evaluate the out of sample predictability of each model on GSI scores. Our evaluation criterion for each model will be the out of sample R^2 as well as the Mean Squared Error (MSE).

4 Results

In this section, we establish how well did our models predict GFSI score, and what possible shortcomings exist with each model. First, it is important to state that the following models are used to predict Global Food Security Index score, not establish causal inference of the indicators. Coefficient estimates, in our case, are therefore uninformative and are considered irrelevant. Since our intention is to focus on out of sample prediction, we will evaluate the out-of-sample R^2 and mean squared error or each model. Producing the three models, each with varying complexity, illustrates which technique yields the most accurate predictions. Section 4.1 explores the three unique models discussed in section 3, in addition to the possible shortcomings each model faces.

4.1 Model Accuracy and Comparisons

We begin by estimating an ordinary least squares regression (OLS) on the Food Security Indicator covariates listed in Table 1 and Table 2. Also listed in Table 3 are the relevant model evaluation criterion. In the OLS model, we find an out-of-sample R^2 of 0.70, which indicates that 70% of the the variability in predicted GFSI score can be explained through our covariates. In addition, the OLS model has a mean squared error of 6.967. While convenient and simple, this particular model has a low out of sample R^2 and a higher MSE relative to other models.

To create a model with greater predictive power, we consider a ridge regression model, which minimizes $L_{optimal}$ as listed in section 3. The main advantage of ridge regression in this case, we assume collinearity is present in the data. This improves the prediction variance of the model at the cost of producing biased results. We find that the ridge model produces an out of sample R^2 value of 0.86; we also notice that the MSE for this model was 2.701. At the cost of biased estimates, the ridge model provides a higher out of sample R^2 and smaller MSE than the OLS model.

The final model we estimate is a random forest, which captures many of the non-linear relationships Ridge and OLS cannot. As the selected indicators are drivers of food security, that they are correlated in some way. A random forest model runs uncorrelated decision tress and produces a balanced reduction in both bias and variance. In the random forest model, we find an out of sample R^2 of 0.902; the highest value of all our models. We also observe an MSE of 2.04; the smallest value among each model. Following directly from the three models, we find that increasing model complexity does results in better out of sample predictive capability.

The selected food insecurity indicators demonstrates a trend of accurate predictions of Global Food Security Index score, with a random forest being the most accurate of the three. In addition to having the highest R^2 value and lowest MSE value, the random forest also displays the most relevant indicators. According to our model, the second lag of average dietary energy supply adequacy (3 year average) is the indicator with the highest importance in predicted GFSI score.Feature selection like that performed on our random forest, is an important application determining the key indicators of food security. Models and studies that examine the most relevant food security indicators are critical in helping us understand how aid is to be distributed.

Although we found strong predictive capability among our three models, we have a number of concerns with these findings. We found a lack of available data for many potential indicators of food security and made the decision to omit some indicators from our models. GFSI data was also limited−GFSI scores have been recorded from 2012 to 2020, which further restricted our data and analysis. Lastly, our model predicts GFSI scores rather than the levels food security− though GFSI can be argued as an adequate and widely used index for identifying levels of food security across countries.

5 Discussion and Conclusion

In this paper, we developed predictive models using food security indicators and GFSI data in LAC countries to provide further insight and solution towards accurately predicting food security across countries. Data from the from the GFSI, Food and Agriculture Organization (FAO), and the World Bank were adequate for applying models to test prediction power. Many food security indices complement their models with dozens of indicators and expert input; we reduced the number of indicators needed to compute and predict GFSI scores and removed qualitative analysis from the evaluation process.

Using a subset of indicators from the World Bank and the Food Agriculture Organization, we constructed a set of 3 models (each of varying complexity) that predicted GFSI scores.

In comparing these computational models—ordinary least squares, ridge, and random forest regression—the ordinary least squares model results in a low out of sample of R^2 and high mean squared error compared to the other models. The random forest model was the best model in predicting GFSI score with an out of sample $R^2 = 0.894$, perhaps indicating there were non-linear relationships captured in the random forest that went undetected in the OLS model. The random forest model also produces a relatively smaller mean squared error of 2.04 compared to the other models.

There are several caveats to our findings, the most significant of them being a severe lack of data. Some of the indicators used in our models reported data on a yearly basis, while overs were based on surveys and reported data every 4-6 years. This makes changes in vital food security indicators very hard to track, and even more difficult to model. Many indicators were of high interest for modeling purposes, but were omitted due to poor data availability. Additionally, our outcome variable of choice (GFSI) has only been measured from 2012-present, meaning our yearly observational units were further restricted to that time frame. We applied imputation to to fill in missing values in our data—all missing values were filled in with the mean indicator value on a country-by-country basis. This method is very simple, but can be costly in model estimation. Because each missing value is assigned the same imputation value, the distribution of the variable becomes skewed and variance is therefore underestimated.

Our work gives light to a few follow up questions, both analytical and administrative. First, our work highlights one of the major problems that food security analysts face: a lack of data to use in modeling and decision making. While it is impossible to collect missing data from the past, quality data collection is imperative moving forward. Investing in high quality data collection will only lead to improved domestic and international response to food insecurity. Another important dimension of understanding food security is correlation between drivers. Modeling the interplay of important food security drivers could show how or why countries may struggle to become food secure. The importance of feature selection must also be mentioned, as understanding relevant food insecurity indicators is paramount in being able to direct relief and aid. This study is important helping governments and nonprofits in their endeavor to provide timely aid to food insecure regions.

Attribution: In the abstract section, Amy wrote the first draft outline; David wrote second outline and proofread; David wrote paragraphs, edited, and proofread. In the introduction section, Amy wrote the outline and paragraphs; David edited paragraphs and proofread. In the literature section, Charlie wrote the outline; Charlie and David wrote paragraphs; Amy and David proofread. In the data section, Nathaniel created tables, wrote/compiled the readme file, and wrote outline and paragraphs; Amy and David added to writing and proofread. In the problem description section, Charlie and David wrote outline; Charlie wrote paragraphs. Amy added to writing and added equations; David proofread. In the results section, Charlie and David wrote outline; Charlie wrote paragraphs; Nathaniel computed regression models and created tables; David proofread. In the conclusion section, Amy wrote outline; Charlie wrote paragraphs; Amy, Charlie, and David added to writing; Amy and David proofread.

References

Alpízar, F., M. Saborío-RodrÃguez, M. Martínez-Rodríguez, B. Viguera, R. Vignola, T. Capitán, and C. A. Harvey (2020). Determinants of food insecurity among smallholder farmer households in Central America: recurrent versus extreme weather-driven events. *Regional Environmental Change 20*(1).

Backer, D. and T. Billing (2021). Validating Famine Early Warning Systems Network projections of food security in Africa, 2009–2020. *Global Food Security 29*, 100510.

Cai, J., J. Luo, S. Wang, and S. Yang (2018). Feature selection in machine learning: A new perspective. *Neurocomputing 300*, 70–79.

Corral, L., G. Gordillo, and P. C. Winters (2000). *Food insecurity and vulnerability in Latin America and the Caribbean.* University of New England, Graduate School of Agricultural and Resource Economics.

EIU (2019). The global food security index 2019: Strengthening food systems and the environment through innovation and investment. Economist Intelligence Unit.

FAO (1996). Rome declaration on world food security and World Food Summit plan of action: World Food Summit. Technical Report (no. WFS 96/REP), Rome: Food and Agriculture Organization of the United Nations.

Frongillo, E. A., H. T. Nguyen, M. D. Smith, and A. Coleman-Jensen (2017). Food insecurity is associated with subjective well-being among individuals from 138 countries in the 2014 gallup world poll. *Journal of Nutrition 147*(4), 680–687.

Gbegbelegbe, S., U. Chung, B. Shiferaw, S. Msangi, and K. Tesfaye (2014, aug). Quantifying the impact of weather extremes on global food security: A spatial bio-economic approach. *Weather and Climate Extremes 4*, 96–108.

Henneberry, S. R. and C. P. Carrasco (2014). Global Food Security: Emerging Economies and Diverging Food Markets. *Journal of Food Distribution Research 45*(1), 29–35.

Izraelov, M. and J. Silber (2019). An Assessment of the Global Food Security Index. *Food Security 11*(5), 1135–1152.

Jones, A. D., F. M. Ngure, G. Pelto, and S. L. Young (2013). What are we assessing when we measure food security? A compendium and review of current metrics. *Advances in Nutrition 4*(5), 481–505.

Masih, J., A. Sharma, L. Patel, and S. Gade (2017). Indicators of Food Security in Various Economies of World. *Journal of Agricultural Science 9*(3), 254.

Mbukwa, J. (2013). A model for predicting food security status among households in developing countries. *International Journal of Development and Sustainability 2*(2), 544–555.

Mc Carthy, U., I. Uysal, R. Badia-Melis, S. Mercier, C. O'Donnell, and A. Ktenioudaki (2018). Global food security â Issues, challenges and technological solutions. *Trends in Food Science and Technology 77*(May), 11–20.

Razzaq, A., U. I. Ahmed, S. Hashim, A. Hussain, S. Qadri, S. Ullah, A. Nawaz Shah, A. Imran, and A. Asghar (2021). An Automatic Determining Food Security Status: Machine Learning based Analysis of Household Survey Data. *International Journal of Food Properties 24*(1), 726–736.

Salazar, L. and G. Muñoz (2014). Ensuring food security in LAC in the Context of Covid-19: Challenges and Interventions.

Smith, M. D., W. Kassa, and P. Winters (2017, aug). Assessing food insecurity in Latin America and the Caribbean using FAO's Food Insecurity Experience Scale. *Food Policy 71*, 48–61.

Thomas, A., B. D'Hombres, C. Casubolo, K. F, and S. M (2017). The use of the Global Food Security Index to inform the situation in food insecure countries. Technical report, EUR 28885 EN, Publications Office of the European Union, Luxembourg.

Weber, M. T., J. M. Staatz, J. S. Holtzman, E. W. Crawford, and R. H. Bernsten (1988). Informing Food Security Decisions in Africa: Empirical Analysis and Policy Dialogue. *American Journal of Agricultural Economics 70*(December), 1044–1052.

Wheeler, T. and J. von Braun (2013). Climate Change Impacts on Global Food Security. *Science 341*(6145), 508– 513.

Table 1: Food Security Indicators by Country
Means and (Standard Errors)
{Argentina, Bolivia, Brazil, Chile, Ecuador}

Food Security Indicator	Argentina	Bolivia	Brazil	Chile	Ecuador
Source - FAO					
Average dietary energy supply adequacy (%) (3-year Average)	131.3	100.9	132.1	121.9	106.7
	(2.31)	(3.17)	(1.29)	(0.88)	(5.29)
Value of crop and livestock exports (in millions of dollars)	36035.85	1445.68	77443.55	10956.67	5305.41
	(4011.68)	(304.20)	(6465.75)	(939.63)	(665.56)
Value of crop and livestock imports (in millions of dollars)	2321.79	681.92	9861.30	5772.40	1957.61
	(913.56)	(113.79)	(959.61)	(763.43)	(227.71)
Total agricultural emissions (in CO_2 equivalent teragrams)	111.61	24.03	447.72	10.48	12.60
	(4.52)	(0.68)	(6.56)	(0.82)	(1.08)
Source - World Bank	(1)	(2)	(3)	(4)	(5)
Personal remittances, received (% of GDP)	0.103	3.856	0.137	0.025	2.854
	(0.026)	(0.482)	(0.024)	(0.002)	(0.423)
Agriculture, forestry, and fishing, value added (annual % growth)	5.131	4.260	3.527	1.892	3.492
	(16.265)	(2.419)	(5.592)	(4.173)	(3.118)
Agriculture, forestry, and fishing, value added (% of GDP)	6.072	10.638	4.413	3.639	9.212
	(0.720)	(0.902)	(0.222)	(0.224)	(0.370)
Adjusted net national income per capita (annual % growth)	0.820	2.894	0.419	3.021	0.916
	(5.711)	(0.987)	(2.959)	(3.517)	(3.844)
Official exchange rate (LCU per US$, period average)	14.29	6.92	2.75	587.01	1.00
	(14.108)	(0.034)	(0.857)	(87.028)	(0)
Employment in agriculture - female (% of female employment)	0.156	29.830	5.232	4.917	22.803
	(0.133)	(1.309)	(1.006)	(0.286)	(2.513)
Employment in agriculture (% of employment)	0.431	29.728	10.577	9.604	27.509
	(0.386)	(1.149)	(1.206)	(0.513)	(1.293)
Food imports (% of merchandise imports)	3.601	7.666	5.427	8.629	9.113
	(1.733)	(0.492)	(0.938)	(1.178)	(1.458)
Food exports (% of merchandise exports)	55.854	15.342	33.817	21.989	39.407
	(4.246)	(2.138)	(2.464)	(3.600)	(9.685)

Table 2: Food Security Indicators by Country
Means and (Standard Errors)
{El Salvador, Guatemala, Honduras, Nicaragua, Peru}

Food Security Indicator	El Salvador	Guatemala	Honduras	Nicaragua	Peru
Source - FAO					
Average dietary energy supply adequacy (%) (3-year Average)	114.1	112.9	110.0	111.5	116.2
	(0.99)	(1.66)	(3.71)	(1.72)	(2.49)
Value of crop and livestock exports (in millions of dollars)	1020.91	4844.30	2180.86	1872.86	5163.97
	(111.50)	(510.35)	(311.09)	(268.10)	(1196.37)
Value of crop and livestock imports (in millions of dollars)	1793.88	2482.98	1502.63	9533.19	4301.07
	(196.09)	(332.03)	(166.63)	(114.35)	(479.08)
Total agricultural emissions (in CO_2 equivalent teragrams)	2.51	8.76	6.10	9.2	23.36
	(0.24)	(0.40)	(0.23)	(0.79)	(0.27)
Source - World Bank	(1)	(2)	(3)	(4)	(5)
Personal remittances, received (% of GDP)	19.019	10.962	17.732	10.189	1.470
	(1.02)	(1.46)	(1.83)	(1.32)	(0.11)
Agriculture, forestry, and fishing, value added (annual % growth)	0.200	2.887	4.462	2.367	3.157
	(5.08)	(1.65)	(3.73)	(3.30)	(3.31)
Agriculture, forestry, and fishing, value added (% of GDP)	5.913	10.123	12.336	16.339	6.874
	(0.78)	(0.62)	(1.00)	(1.27)	(0.14)
Adjusted net national income per capita (annual % growth)	1.411	2.095	1.622	2.104	3.887
	(2.18)	(2.05)	(3.31)	(4.47)	(3.59)
Official exchange rate (LCU per US$, period average)	1.00	7.71	21.53	26.86	3.89
	(0)	(0.20)	(2.10)	(3.96)	(0.29)
Employment in agriculture - female (% of female employment)	4.457	10.871	9.346	8.387	26.563
	(0.99)	(1.53)	(1.38)	(0.65)	(0.43)
Employment in agriculture (% of employment)	19.007	31.716	32.033	30.477	27.838
	(1.76)	(1.12)	(3.54)	(0.55)	(0.38)
Food imports (% of merchandise imports)	16.829	14.047	17.811	15.698	10.860
	(0.86)	(0.83)	(1.00)	(1.35)	(0.78)
Food exports (% of merchandise exports)	20.224	44.366	60.005	52.654	18.819
	(2.19)	(2.61)	(5.51)	(12.66)	(2.73)

Table 3: GFSI Score by Country

Country	2012	2013	2014	2015	2016	2017	2018	2019	2020
Argentina	61.4	62.7	62.1	61.3	64	63.3	65.7	65.0	62.7
Bolivia	54.7	54.4	57.5	62.1	61.0	61.4	61.9	60.6	60.0
Brazil	65.1	66.4	65.8	66.6	69.5	71.7	69.5	66.6	64.1
Chile	68.5	68.8	68.4	69.0	69.1	71.2	72.8	73.0	70.2
Ecuador	57.6	58.3	57.5	59.7	60.4	60.9	60.7	59.9	57.9
El Salvador	59.1	59.3	58.9	61.2	59.6	58.1	60.5	60.6	59.0
Guatemala	52.4	53.0	57.5	53.5	56.9	55.1	57.8	56.5	56.2
Honduras	53.7	54.8	56.9	59.5	57.3	56.5	58.9	58.7	58.2
Nicaragua	52.5	53.2	53.1	58.2	58.4	55.5	54.8	53.5	54.4
Peru	57.6	60.8	59.3	62.2	63.1	64.2	63.1	65.3	65.7

The distribution of all GFSI scores has the following notable percentiles:
25^{th} percentile - 49.4; 50^{th} percentile - 62.0; 75^{th} percentile - 72.1; 100^{th} percentile - 85.3
High performing countries for comparison: United States - 77.5, Canada - 77.2, United Kingdom - 78.5
Low performing countries for comparison: Yemen - 35.7, Rwanda - 38.8, Syria - 40.0

Table 4: Model Results
R^2 and Mean Squared Error

Metric	OLS	Ridge	Random Forest
R^2	0.703	0.860	0.902
MSE	6.967	2.701	1.890

Table 5: Top Ten Features from Pruned Random Forest Results
Ranked in Descending Order

Feature	Mean Decrease in Impurity
Average dietary energy supply adequacy (%) (3-Year Average) Lag 2	0.313
Personal remittances, received (% of GDP) Lag 2	0.125
Value of Crop and Livestock Imports (in millions of dollars) Lag 1	0.100
Value of Crop and Livestock Imports (in millions of dollars) Lag 2	0.058
Agriculture, forestry, and fishing, value added (% of GDP) Lag 1	0.057
Average dietary energy supply adequacy (%) (3-Year Average) Lag 1	0.051
Official exchange rate (LCU per US$, period average) Lag 2	0.043
Agriculture, forestry, and fishing, value added (% of GDP) Lag 2	0.039
Official exchange rate (LCU per US$, period average) Lag 1	0.035
Value of Crop and Livestock Exports (in millions of dollars) Lag 1	0.033

Lag 1 represents the feature data are collected one year prior. Lag 2 represents the data are collected 2 years prior.