Using Machine Learning Models to Analyze Standardized Agricultural Survey Data Collected in 21 Countries

Authors: Ana Boeriu, Christina De Cesaris, Mary-Francis LaPorte

I. Abstract

RHoMIS, a massive standardized survey effort in global agriculture, has allowed scientists to consider the association between household factors and poverty/farm productivity. Before now, predictive feature selection methods have not been used to study a 21-country survey of household farm productivity. In this study, we fit parametric (linear regression, ridge regression, LASSO) and non-parametric (Random Forest) machine learning models to analyze indicator variables. We find that geography are good predictors of Market Orientation (MO) and Percent Poverty Index (PPI), but the top specific predictors change in each model. Furthermore Random Forest is the best model overall. These findings open the door to the applications of predictive testing on agricultural survey data.

II. Introduction

RHoMIS survey procedures allow for international agriculture researchers to apply standardized methodologies to generate a globally comparable picture of household and farm demographics. To this date, machine learning methodologies have not been fit to predict indicator values based on survey data. Applying predictive models may allow scientists to determine the reason behind trends identified by the machine learning models. Additionally, once a model is deemed properly fit and versatile for RHoMIS datasets, machine learning could be used to aid in farm consulting.

The data used in this study was previously published in an article entitled “The Rural Household Multiple Indicator Survey (RHoMIS) data of 13,310 farm households in 21 countries”, in *Scientific Data* (a Nature publication) in 2019 (doi) The data is all publicly accessible through the Harvard Dataverse. In the experiment, the researchers surveyed over 13,000 families in 21 countries using a standardized survey approach. By standardizing both the questions, possible responses, and time to survey, they assert that the data reported by each household is advantageously reliable, compared to other surveying methods. They collected dozens of variables about each family, and calculated around 50 indicators based on the collected information. In this study, we utilize a selection of the raw variables and indicators (that are not confounding with the response variables of PPI and MO). For a full list of the variables used in our study, see the clean.csv dataset under the data directory in the project github repository. For a full description of all variables and their meaning, see “Raw data code book.tab” and “Explanation_of_Calculations_and_Outputs.tab” from the Harvard Dataverse public data (doi). Any variables significant in our study will be explained in the report.

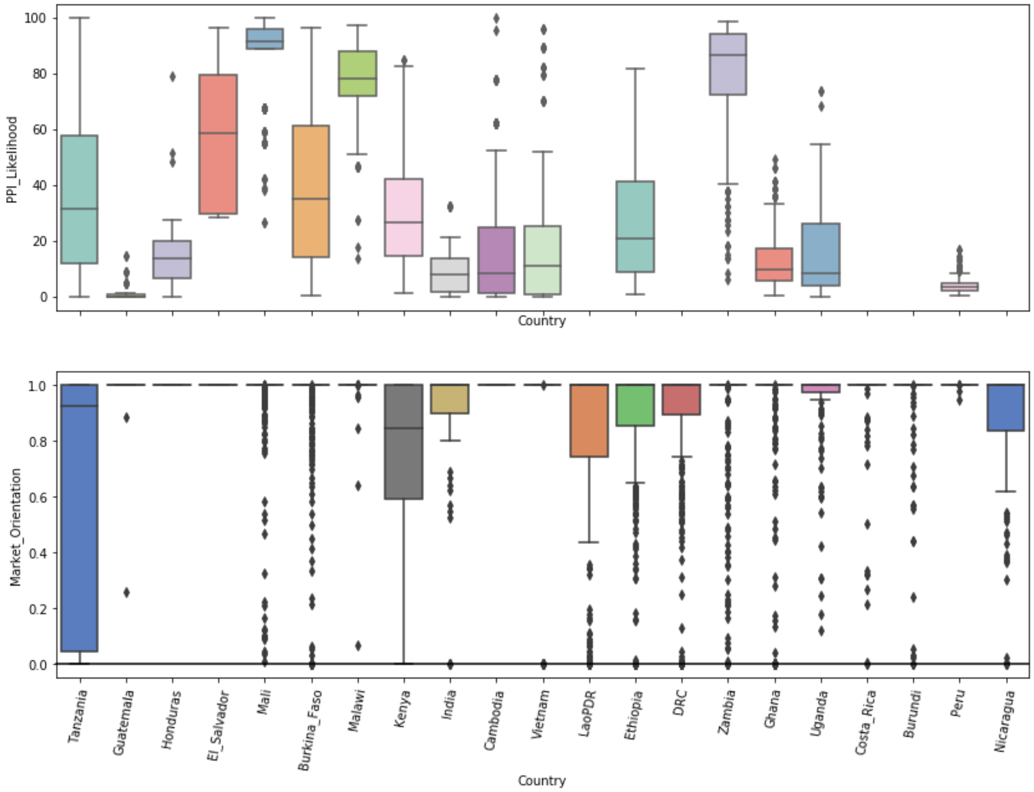
The main research question for this project is: “how will different machine learning models perform in predicting the PPI and MO based on other survey factors?” Furthermore, we aim to identify the most important coefficients in predicting PPI and MO. PPI is a widely-used indicator to determine if a household lives in what is considered poverty. The term PPI is standardized in meaning across disciplines because it is actually a registered trademark (Poverty Probability Index®) of the PPI alliance. The indicator takes the answers to a list of pre-determined questions that are known to be highly correlated to living in poverty, and thereby calculates the probability that the family being surveyed is in poverty based on the contents of their response. In this study, we did not calculate the PPI-- it was already determined by the authors and collectors of the original experiment and dataset. Market orientation is calculated as the percentage of an agricultural good that has been produced that will be sold at the market. MO can be used to describe both crops and livestock. The MO was also included in the original dataset.

III. Exploratory Data Analysis

Exploratory data analysis focused on identifying country-level trends in the dataset and response variable selection. First, the correlation between the numerical predictors was plotted to gain a better understanding of how different features were linearly related. The correlation plot combined with intuition from the data cleaning process led us to choose PPI_Likelihood and Market_Orientation as the predictors for our models. The correlation plot, although not included in this report due to size, can be found on the github repo under the title /misc/Charts_Graphs_EDA.ipynb

We then produced a box plot of our response variables against the countries still present in the data after cleaning (Figure 1). The boxplot for PPI_Likelihood against countries indicated that there was high variability in response between countries. This variability was not present in the boxplot relating Market Orientation to countries. This exploratory insight presented a potential set back, as it implied that PPI Likelihood was country specific and therefore the models may primarily predict PPI Likelihood using the country's name and not the other features as we intended.

Figure 1: Box Plots: countries vs responses (ppi, mo)



IV. **Data cleaning**: (see STA_208/finalized_code/more_data_cleaning.ipynb)

The first step for data cleaning was to choose the variables to include in the study using brief background information gained from the "Raw data code book" and "Explanation_of_Calculations_and_Outputs.tab". The goal of this was to omit anything that was a) redundant with any of the response variables b) irrelevant to our study scientifically or, rarely, c) where the meaning of the values were unclear. We did this by hand, as a group, manually discussing each of the ~200 variables, and reducing them to the ~40 to be used in the final model. The precleaned files are included on the github repo and entitled "RHoMIS_Full_Data.csv" and "RHoMIS_Indicators.csv". They contain the raw variables and non-confounding indicators, respectively.

Upon further data exploration, we found a significant number of NA's. Variables that had more than 30 percent of NAs were removed since imputing any amount greater than that would greatly skew our results. For the numeric variables we imputed using the median as it is less biased compared to using the mean. A missingness category of 'na' was created for categorical variables with NA values. We then removed all rows where our response variable was NA. Furthermore, we also adjusted the Education column levels, as described in the next paragraph. All variable names were converted to lowercase, and the variable types (numeric, categorical) were verified and corrected if needed.

Head Education Level (or the highest level of education achieved by the head of the household) was recorded as being either "primary, secondary, post-secondary, literate, illiterate, or no school". In some surveying contexts (i.e. groups of surveys administered by the same person), the survey administrator chose to record non-standard descriptions of the head education level. Examples of such non-standard descriptions were describing the religion of the parochial school but not the education level, recording the education level in the context of the language that was spoken at the school, or using locally-recognized literacy ratings without converting them to the standardized responses. As these non-standard descriptors are not clearly defined, it was impossible to map them to one of the standardized responses. As a result, the rows with non-standard head education levels were dropped. This did not affect the majority of rows for any country, and it was primarily an issue in Burkina Faso and Mali.

Finally, we manually added the continent corresponding to each country in the cleaning file. In downstream analyses, we used this coding to compare the results. For Market Orientation analysis using linear models, the names of the countries were included as categorical variables, but continent was omitted. For PPI, the name of the continent was included as a categorical variable, but the name of the country was omitted.

V. Model Methods

Given the complexity of the data, we decided to train both parametric and nonparametric models and evaluate feature importance accordingly. The optimally tuned linear parametric models--OLS, Ridge, and LASSO regression--tended to have higher mean squared error values than their nonparametric Random Forest Regressor counterpart (see Table 5 in the appendix). The Random Forest feature importances are based on the mean decrease in impurity which occurs when a tree splits on a specific feature. The feature importance scores are not indicative of how a particular feature relates to the response variable, but rather its significance in determining the splits across all estimators in the Random Forest. For the parametric models, feature importance was extracted from the values of the model coefficients which allowed us to investigate the relationship between the response variables and predictors.

The linear models were trained using the cleaned dataset and either the PPI_Likelihood or Market_Orientation variables as the response. It was often the case that the dominating features were those representing the different countries present in our dataset. For example, in the PPI_Likelihood models, the countries in Africa would overshadow the importance of other features. This issue was handled by replacing the country column with the continent to which each country belonged. Unsurprisingly, the continent Africa was found a significant predictor in the new models, but other features which were previously subdued emerged.

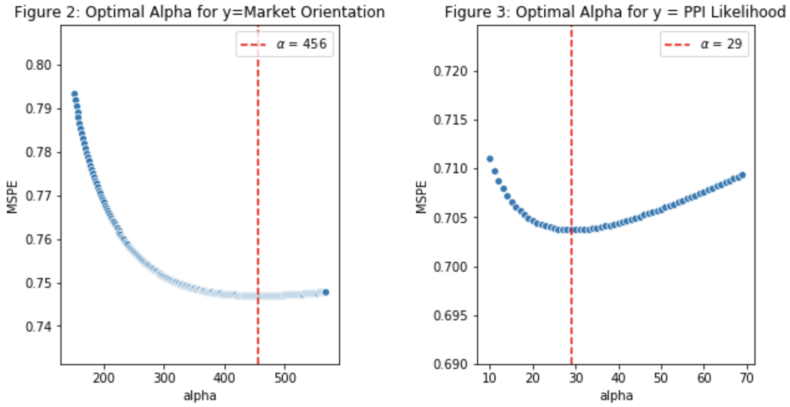
Ordinary Least Squares was performed on the cleaned dataset. The analysis was run first with just the numerical data, then including the categorical variables as dummy variables. The MSE values for OLS are reported in Table 5. Standardization was completed using a custom function that performed a z-score transformation. The results were reported using the statsmodels.api library. These results are included in the LinearModel.ipynb notebook.

Separate, country-specific Random Forest models were run on subsets of the data to gain a better understanding of which features were the most influential in predicting the PPI_Likelihood for each country. Continent-specific models were also run for comparison. The Random Forest Regressor was chosen for the subsetted data because the Random Forest model performed best on the dataset that is not subsetted by country.

VI. Cross Validation

Before we are able to fit the respective models we performed cross validation to obtain the optimal parameters based on the criterion of mean square prediction error. This allows us to optimize the performance of each model and prevent overfitting the data. All models are fit without intercept since we do not have baseline categories (Figure 2).

Figure 2: Ridge Regression



Ridge regression decreases model complexity and reduces the variance by using an additional regularization parameter ,α, that controls the size of our beta coefficients. Using leave one out cross validation, as our α increases, the smaller our coefficients will be. From Figures 3 and 4 we can see that the using Market orientation as our response requires a larger α than using PPI as a response, which could underfit the model.

A. LASSO

The LASSO Regression hyperparameter tuning was performed using 5 fold cross validation and plotting the average MSE from all folds given the hyperparameter α. Overall the α values for both the PPI and MO models was quite small and large values of α significantly increased the associated MSE

B. Random Forest

The Random Forest Regression models were tuned using grid search methods with 5 folds on a parameter dictionary containing an array of values for max_depth and n_estimators. Due to the time complexity of the tuning process, only two hyperparameters were chosen for tuning.

VII. Results

In-text results summarized in Tables 1-5 in the Appendix below.

A. OLS Results:

The ordinary least squares model was calculated on numeric values from the dataset only. This is because the MSE was over 10 orders of magnitude smaller when fitting with just the numeric values. These findings were replicated by fitting a ridge regression with no ridges, confirming that this was not an error with data processing. This model had the worst MSE out of all of the models. This is because the linear prediction is not sufficient for this dataset. The variables that were the most significant for OLS were not the same as the variables found as significant for the other three models. The importance of gender control (whether male or female) was the same. This is likely because the values were standardized, and female control is calculated as (1- percentage male control). To further understand these variables, a different processing procedure would be needed. Livestock production ended up being an important predictor of market orientation. This is likely due to the fact that the value of a particular type of cattle is correlated to its desirability in a market. Food availability and the contents of the diet were important predictors of PPI.

B. Ridge Regression Results:

Ridge Regression did significantly better than ordinary least squares regression but not as well as LASSO or Random forest regression (Table 5). Country, year, livestock holdings, crop and household types were all important predictors for market orientation (Tables 1, 2). This makes sense since MO is dependent on the crop season (year) and geographic location of the farmer (country). Similarly for PPI as expected, continent year, education level, and number of household members are all important predictors. Not surprisingly, people with a small edu level, who live in africa will tend to have a high PPI compared to others.

C. LASSO Results:

The LASSO Regression model performed slightly better than the Ridge Regression model when predicting both MO and PPI (Table 5). It produced similar results to the Ridge model regarding the predictors associated with the highest coefficient values. Many of the predictors deemed significant by the Ridge model concerning MO overlapped with those of the LASSO model, and those which did not overlap tended frequently belonged to the same category with the exception of Male Gender Control which was found important in the LASSO MO model but not the Ridge MO model. Regarding the PPI model, the LASSO regression predictors consisted of a mix of significant predictors from the Random Forest model and the Ridge Model. A full table of predictors and MSE values for the models can be found in Table 5.

D. Random Forest Results:

The Random Forest model performed optimally for both the PPI and MO models compared to the other models. As well, the MO model did not find countries to be a significant predictor, but rather livestock production and consumption, farm incomes, and specific crops were found more important. In the case of the PPI model, belonging to the continent of Africa remained the most important predictor followed by the amount of months wild food consumed by families (Table 3, 4; Figure 4,5).

E. Country-Level and Contient-Level Feature Importance, PPI Only

In general, the country- and continent-specific Random Forest models out performed the original model in terms of MSE (Table 6 and 7). However, it is important to recognize that the subsetted data used in the specific models varies drastically across countries. In some cases, the feature importances extracted from the individual models were plotted accordingly, (Figure 4,5) and it was found that many of the predictors deemed significant in the other models made an appearance. The country specific models as well gave insights to why certain predictors may have been selected in the models fitted on the complete dataset. For example, the overwhelmingly significant feature for predicting PPI in Ethiopia was the year 2017, which also appeared to be significant in other countries throughout Africa. However, year 2017 was not highly significant in predicting PPI in many of the non-African countries which implies that in the complete dataset models, the presence of year 2017 is primarily related to harsh conditions in Africa.

VIII. Discussion

To see how much PPI is affected by each predictor, we can look at the ridge regression model
Our ridge regression model has the form

$$Y_{PPI} = 0.02179X_{crop,count} + 0.3866X_{HH,members} - 0.00037X_{land\ owned} + \cdots + 0.6828X_{Africa} - 0.92944X_{central\ am}$$

Generally when we interpret categorical variables, we always compare the current factor level to the baseline factor level. However, because our model does not have baseline categorical levels, our interpretation will change slightly. In our model shown above we are interested in how much the PPI changes for those living in Africa versus those in Central America. We look at the difference between the two sub models for Africa and Central America. Thus, comparing Central America to Africa, the PPI would decrease by 1.6122 units on average, holding all other variables constant.
A general interpretation of a numeric variable is when X_1 increases by 1 unit, Y increases by β_1 units on average holding all other variables constant. Thus interpreting the variable land owned: When a person buys a hectare of land, the PPI decreases by 0.00037 units on average holding all other variables constant.

We treated the variable years as categorical variables. Conceptually similar to the geographic separation of countries being a categorical variable, the temporal separation of each year was treated as its own environment with myriad factors having the potential to change. One large assumption that this treatment requires is that the survey answers are independent between years (to fulfil the independence assumption). This means that we are assuming that the families being surveyed in one year are distinct from those being surveyed the next. Alternatively, this could be true if the conditions were independent enough from year to year in one family, that the assumption could be met. We are unable to know if the families surveyed are consistent over the years because the survey data is de-identified. In summary: we assume sample independence.

The type of crop ended up being a major predictor in almost every case-- but the individual crops identified were not major global food crops. This suggests that the crops being significant were ones grown in very few households. This means that the significance may be in error, from the small sample sizes for these crops. Figure 3 in the appendix demonstrates the frequency of crop types on farms to depict this trend. In future analyses with this data, the crop frequency should be taken into account. One method for doing this would be to only include those which more than 20 farmers grew. Another factor to consider is the geographic distribution of crops. The crops might be a confounding predictor with the country, so further analysis needs to be done.

IX. Conclusion

We found that Random Forest was the most accurate machine learning model to analyze these experimental results. These prediction results could help direct research questions to explore the relationship between PPI or MO and these variables. For example, this could be used to explore if there is a causal relationship between livestock ownership and MO, or if they are confounded by another factor (for example, the ability to own livestock). These results are not fine-tuned enough to use for consulting, but this preliminary set of results could aid in that direction. Future experiments on this data should consider “by year” interactions. In particular, it would be beneficial to survey the same family multiple years in a row, to see the variance in the same family between years.

Link to Github Repo: https://github.com/laporpe/STA_208

Reference:

van Wijk, M., Hammond, J., Gorman, L. *et al.* The Rural Household Multiple Indicator Survey, data from 13,310 farm households in 21 countries. *Sci Data* 7, 46 (2020). https://doi.org/10.1038/s41597-020-0388-8

Appendix:

Table 1: Top 9 Coefficients (or Feature Importance Score for Random Forest) for Market Orientation for each model. Note: Country included, not continent. OLS was calculated on numeric values only.

	OLS		Ridge Regression		LASSO		Random Forest	
1	Livestock production	-2.212e+13	Country, Tanzania	-0.4320	Country Tanzania	-0.7046	Value livestock produced and consumed	0.4296
2	Value of livestock	2.014e+13	Country, Kenya	-0.2481	Country Kenya	-0.4281	Farm income	0.3309
3	Gender control male	1.566e+13	Year, 2017	-0.2245	Year 2017	-0.3615	value_crop_consumed	0.1213
4	Gender control female	1.566e+13	Country, Burkina Faso	0.1891	Livestock Holdings	-0.1435	Total income	0.1017
5	Livestock product sales	9.126e+12	Livestock Holdings	-0.1817	Country Mali	0.1418	Value farm produce	0.0044
6	Value farm produce	9.066e+11	Country, Mali	0.1458	Gender Male Control	-0.1359	Crop sales	0.0023
7	Crop sales	-6.911e+11	Country, Laos	-0.1160	Country Ethiopia	-0.1331	Livestock product sales	0.0014
8	Total income	-2.066e+11	Crop, Greengram	0.1073	Country Burkina Faso	0.1159	TVA (in USD)	0.0007
9	Off-farm income	1.321e+11	Type of Household: Woman, single	0.0927	Crop Harvest na	0.0756	Value livestock production	0.0007
10	Farm income	-9.777e+10	Year, 2016	0.0904	Crop Intercrop, na	0.0736	Value Crop produce	0.0006

Table 2: Top variable categories for Market Orientation (in order of appearance)

	OLS	Ridge Regression	LASSO	Random Forest
1	Livestock Production	Country	Country	Value of Livestock Consumed

2	Gender Control	Year	Year	Income
3	Farm Production	Livestock Holdings	Livestock Holdings	Value of Crops Consumed
4	Crop sales	Crop	Gender Male Control	Crop Sales
5	Total income	Household Type	Crop	Livestock Sales

Table 3: Top 9 Coefficients (or Feature Importance Score for Random Forest) for PPI for each model. Note: Continent included, not country. OLS was calculated on numeric values only.

	OLS		**Ridge Regression**		**LASSO**		**Random Forest**	
1	Food availability	0.8301	Continent, Central america	-0.9359	Continent Africa	0.6607	Continent, Africa	0.0931
2	Household size	0.82577	Head Education Level, none	-0.8022	Year, 2015	0.4607	Number of Months Wild Food Consumption	0.0713
3	Total Value of daily activities	-0.7412	Continent, Africa	0.6756	Head Education Level, No school	0.3961	Livestock Holdings	0.0538
4	Household size (normalized)	-0.7160	Head Education level, no school	0.5045	Year 2017	0.3572	Household size male adult equiv	0.0519
5	Off-farm income	-0.1507	Year 2015	0.4944	Number of Months, Wild food consumption	0.124081	Year 2017	0.0449
6	Number of months food consuming wild food	0.1013	Year, 2018	-0.4888	Crop, millet	0.0871742	Food Self Sufficiency kCal, MAE, day	0.0435
7	Number of months food	0.0973	Crop, irish potato	-0.4335	Head Education	0.0814309	Number of Household	0.0382

	insecure				Level, primary		members	
8	Dietary diversity during a bad season	-0.0683	Number of household members	0.3781	Number of Household members	0.0690304	HDDS score, good season	0.0373
9	Crop count	0.0670	Crop, teff	-0.3764	HouseholdType_couple	0.0597672	Year 2015	0.0325
10	Total income	-0.0584	Crop, wheat	-0.334	HFIAS_status_severelyfi	0.0566652	HDDS score, bad season	0.0324

Table 4: Top variable categories for PPI (in order of appearance)

	OLS	**Ridge Regression**	**LASSO**	**Random Forest**
1	Food availability	Continent	Continent	Continent
2	Household size	Head Education level	Year	Number of Months Wild Food was Consumed
3	Value of daily activities	Year	Head Education Level	Livestock Holdings
4	Off-farm income	Crop	Number of Months Wild Food was Consumed	Household Size Caloric Intake
5	Food insecurity/ Dietary diversity	Number of household members	Crop	Year

Table 5: Model evaluation criteria. Mean Squared Error. PPI uses only continent and not country, MO uses country but not continent.

	OLS	**Ridge Regression**	**LASSO**	**Random Forest**

	PPI	MO	PPI	MO	PPI	MO	PPI	MO
MSE	0.9815	3.2187	0.8451	0.7104	0.6969	0.6696	0.4778	0.0099
alpha			29	456	0.01	0.01		

Table 6: Random Forest Country Specific PPI Model Mean Squared Error Values

Country	**MSE**
Tanzania	0.4789
Guatemala	0.0059
Honduras	0.361
El Salvador	0.7907
Mali	0.2037
Kenya	0.6274
India	0.3425
Cambodia	0.2476
Vietnam	0.0499
Ethiopia	0.3154
Zambia	0.2916
Ghana	0.1657

Table 7: Random Forest Contient Specific PPI Model Mean Squared Error Values

Contient	**MSE**
Africa	0.5036
Central America	0.1999
South America	0.0142
Asia	0.2853

Figure 3. The frequencies of crop (crop_name_1) for all crops grown in more than 20 households. Any crops grown in less than 20 households are omitted (not to be confused with “other”).

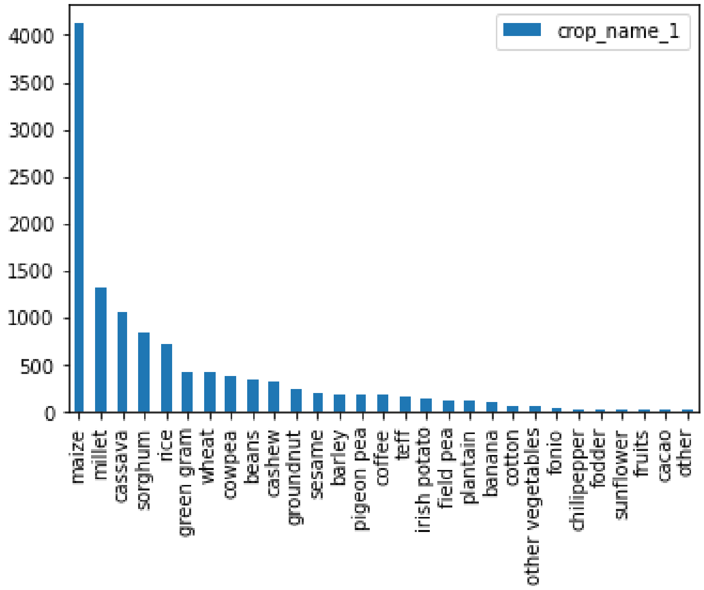
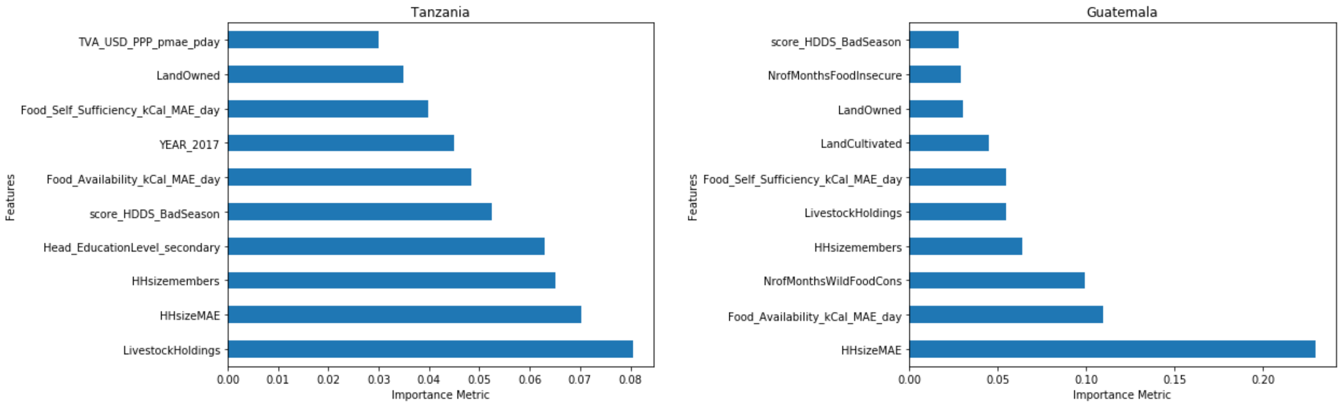
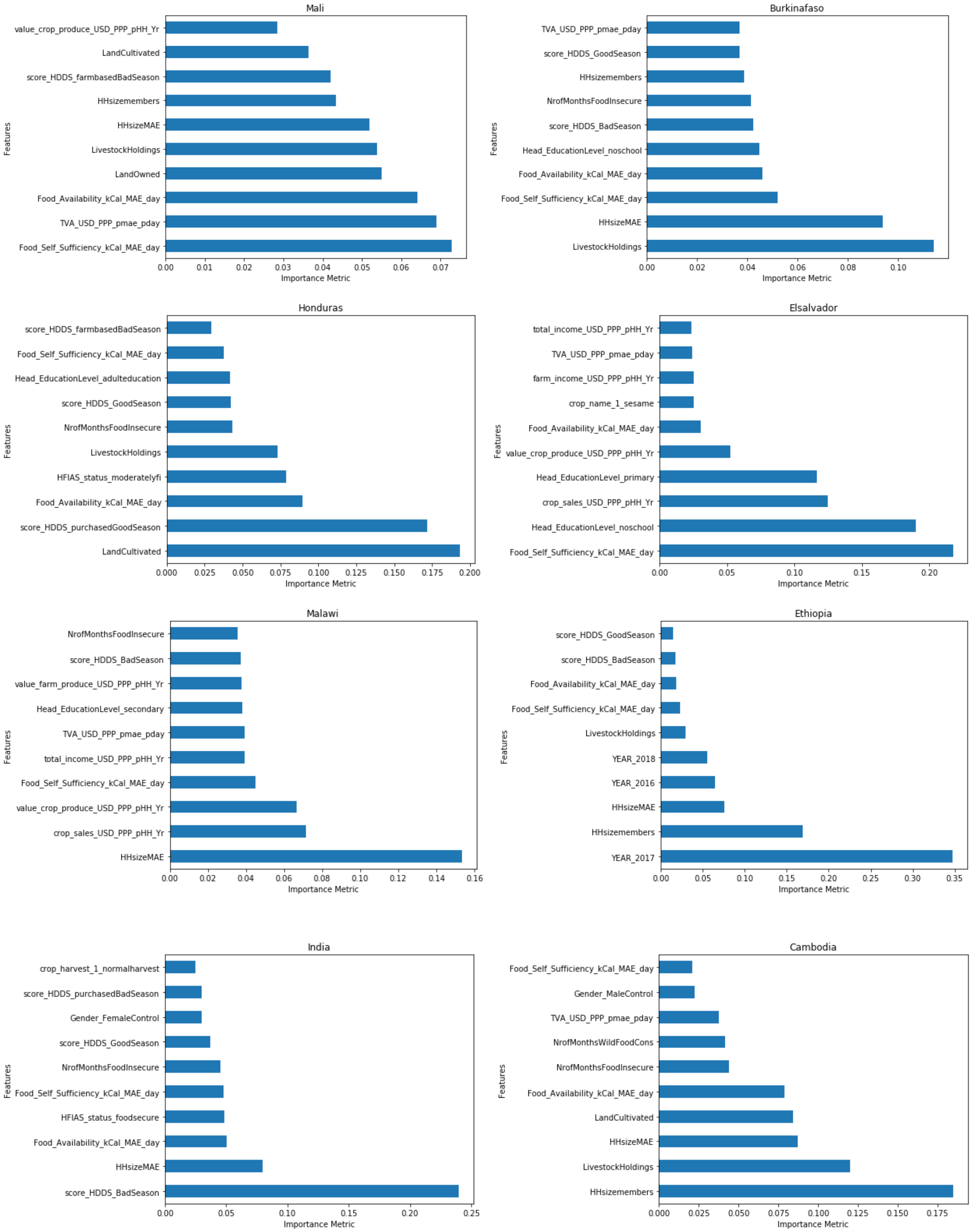


Figure 4: PPI




Mali
value_crop_produce_USD_PPP_pHH_Yr
LandCultivated
score_HDDS_farmbasedBadSeason
HHsizemembers
HHsizeMAE
LivestockHoldings
LandOwned
Food_Availability_kCal_MAE_day
TVA_USD_PPP_pmae_pday
Food_Self_Sufficiency_kCal_MAE_day
Features
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07
Importance Metric
Burkinafaso
TVA_USD_PPP_pmae_pday
score_HDDS_GoodSeason
HHsizemembers
NrofMonthsFoodInsecure
score_HDDS_BadSeason
Head_EducationLevel_noschool
Food_Availability_kCal_MAE_day
Food_Self_Sufficiency_kCal_MAE_day
HHsizeMAE
LivestockHoldings
Features
0.00 0.02 0.04 0.06 0.08 0.10
Importance Metric
Honduras
score_HDDS_farmbasedBadSeason
Food_Self_Sufficiency_kCal_MAE_day
Head_EducationLevel_adulteducation
score_HDDS_GoodSeason
NrofMonthsFoodInsecure
LivestockHoldings
HFIAS_status_moderatelyfi
Food_Availability_kCal_MAE_day
score_HDDS_purchasedGoodSeason
LandCultivated
Features
0.000 0.025 0.050 0.075 0.100 0.125 0.150 0.175 0.200
Importance Metric
Elsalvador
total_income_USD_PPP_pHH_Yr
TVA_USD_PPP_pmae_pday
farm_income_USD_PPP_pHH_Yr
crop_name_1_sesame
Food_Availability_kCal_MAE_day
value_crop_produce_USD_PPP_pHH_Yr
Head_EducationLevel_primary
crop_sales_USD_PPP_pHH_Yr
Head_EducationLevel_noschool
Food_Self_Sufficiency_kCal_MAE_day
Features
0.00 0.05 0.10 0.15 0.20
Importance Metric
Malawi
NrofMonthsFoodInsecure
score_HDDS_BadSeason
value_farm_produce_USD_PPP_pHH_Yr
Head_EducationLevel_secondary
TVA_USD_PPP_pmae_pday
total_income_USD_PPP_pHH_Yr
Food_Self_Sufficiency_kCal_MAE_day
value_crop_produce_USD_PPP_pHH_Yr
crop_sales_USD_PPP_pHH_Yr
HHsizeMAE
Features
0.00 0.02 0.04 0.06 0.08 0.10 0.12 0.14 0.16
Importance Metric
Ethiopia
score_HDDS_GoodSeason
score_HDDS_BadSeason
Food_Availability_kCal_MAE_day
Food_Self_Sufficiency_kCal_MAE_day
LivestockHoldings
YEAR_2018
YEAR_2016
HHsizeMAE
HHsizemembers
YEAR_2017
Features
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35
Importance Metric
India
crop_harvest_1_normalharvest
score_HDDS_purchasedBadSeason
Gender_FemaleControl
score_HDDS_GoodSeason
NrofMonthsFoodInsecure
Food_Self_Sufficiency_kCal_MAE_day
HFIAS_status_foodsecure
Food_Availability_kCal_MAE_day
HHsizeMAE
score_HDDS_BadSeason
Features
0.00 0.05 0.10 0.15 0.20 0.25
Importance Metric
Cambodia
Food_Self_Sufficiency_kCal_MAE_day
Gender_MaleControl
TVA_USD_PPP_pmae_pday
NrofMonthsWildFoodCons
NrofMonthsFoodInsecure
Food_Availability_kCal_MAE_day
LandCultivated
HHsizeMAE
LivestockHoldings
HHsizemembers
Features
0.000 0.025 0.050 0.075 0.100 0.125 0.150 0.175
Importance Metric

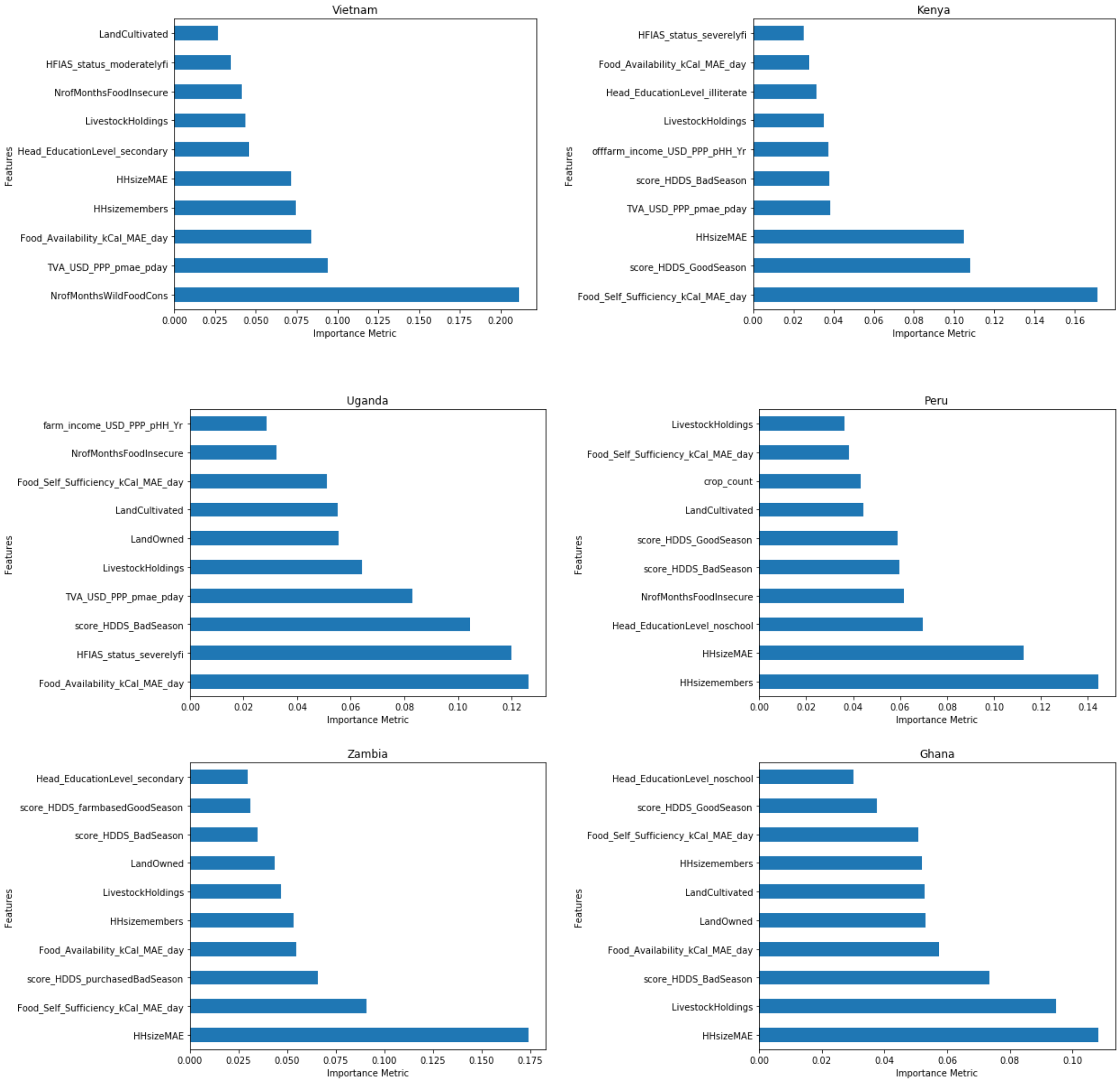
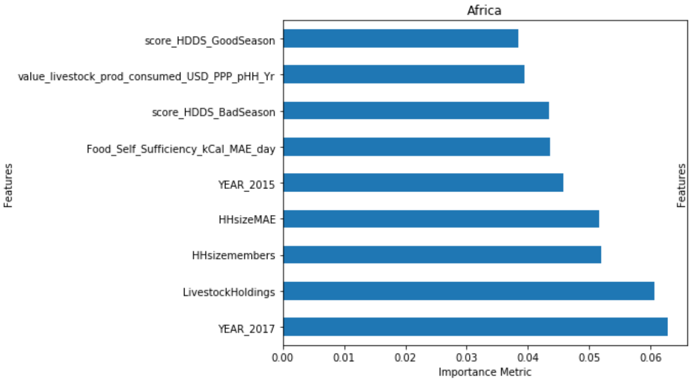
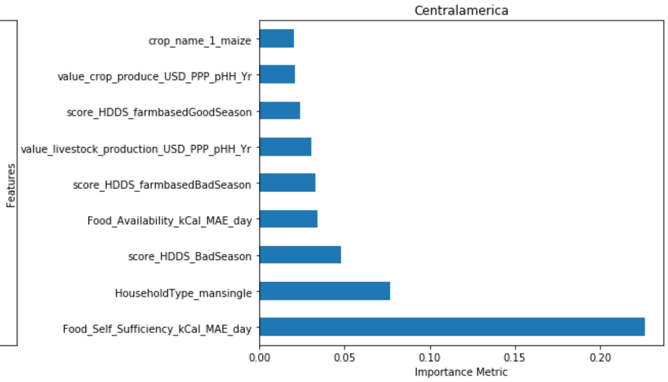
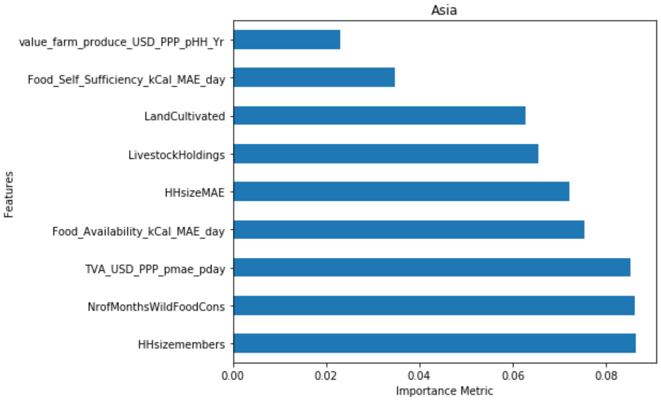
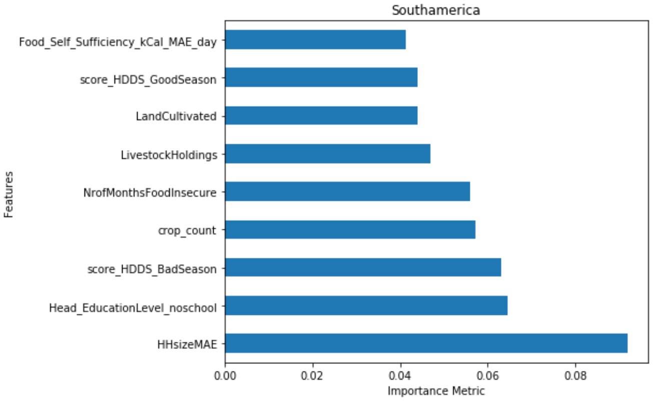


Figure 5: Market Orientation


Africa
score_HDDS_GoodSeason
value_livestock_prod_consumed_USD_PPP_pHH_Yr
score_HDDS_BadSeason
Food_Self_Sufficiency_kCal_MAE_day
YEAR_2015
HHsizeMAE
HHsizemembers
LivestockHoldings
YEAR_2017
Features
0.00
0.01
0.02
0.03
0.04
0.05
0.06
Importance Metric


Centralamerica
crop_name_1_maize
value_crop_produce_USD_PPP_pHH_Yr
score_HDDS_farmbasedGoodSeason
value_livestock_production_USD_PPP_pHH_Yr
score_HDDS_farmbasedBadSeason
Food_Availability_kCal_MAE_day
score_HDDS_BadSeason
HouseholdType_mansingle
Food_Self_Sufficiency_kCal_MAE_day
Features
0.00
0.05
0.10
0.15
0.20
Importance Metric


Asia
value_farm_produce_USD_PPP_pHH_Yr
Food_Self_Sufficiency_kCal_MAE_day
LandCultivated
LivestockHoldings
HHsizeMAE
Food_Availability_kCal_MAE_day
TVA_USD_PPP_pmae_pday
NrofMonthsWildFoodCons
HHsizemembers
Features
0.00
0.02
0.04
0.06
0.08
Importance Metric


Southamerica
Food_Self_Sufficiency_kCal_MAE_day
score_HDDS_GoodSeason
LandCultivated
LivestockHoldings
NrofMonthsFoodInsecure
crop_count
score_HDDS_BadSeason
Head_EducationLevel_noschool
HHsizeMAE
Features
0.00
0.02
0.04
0.06
0.08
Importance Metric