# Using Machine Learning to Improve Treatment Targeting in Farmer Training *

Betsy Alter † Andres Perez Martinez ‡ Joshua Short § Su-Jung Teng ¶

July 12, 2023

Preliminary draft, please do not cite.

**Abstract**

Climate change and food insecurity threaten the livelihoods of smallholder farmers. Our study focuses on predicting the adoption of good agricultural practices (GAP) among farmers, specifically in Cambodia and Ethiopia. We ask: Which machine learning model most accurately predicts which farmers will implement these sustainable practices? We analyze comprehensive survey data, Normalized Difference Vegetation Index (NDVI), and precipitation data. Our Support Vector Machine model achieves the highest accuracy and precision. We also perform cross-country analysis, and we achieve promising precision metrics. Using Random Forest feature importance and Shapley values, we find longitude, latitude, and amount of land used for farming to be the features that drive prediction the most. Despite limitations, these findings highlight the efficacy of machine learning in predicting GAP adoption. Our results can guide policymakers to design targeted agricultural training programs to efficiently use resources and improve the productivity of farmers while reducing their climate change impact.

Keywords: agriculture, machine learning, adoption prediction, cross-country analysis, Good Agricultural Practices (GAP), NDVI

## 1 Introduction

Despite widespread investment in agricultural extension training, few studies aim to predict who will adopt new agricultural practices and make lasting changes (Kansanga et al., 2021; Moser and Barrett, 2003). Predicting adoption is necessary both for allocating resources appropriately for training as well as predicting which households may be more vulnerable if they do not adopt. The difficulty with effective prediction models (as opposed to the more traditional economic objective that focuses on unbiased causal inference of intervention's effects) can lie in the numerous inputs needed in order to predict adoption since the variance in adoption even within a small region can

*We would like to thank the World Bank, Professor Vasilaky, and Cal Poly for this opportunity and help throughout the completion of this project

†(e-mail: balter@calpoly.edu)
‡(e-mail: apere224@calpoly.edu)
§(e-mail: jshort06@calpoly.edu)
¶(e-mail: suteng@calpoly.edu)

vary substantially.

For example, in China, only 40% of farmers receive agricultural training. And of these, 51% end up adopting the techniques that they are trained in (Yang et al., 2021). We know that adopting new agricultural techniques, particularly climate-smart techniques, is increasingly important. According to the Global Commission on Adaption, without implementing GAP, we may see a decrease in global agricultural yields between 5% and 30% (Ki-moon et al., 2019).

Training farmers in what are known as Good Agricultural Practices (GAP) are an effective way to increase long-term yield and combat climate change and soil degradation. Professional, workshop, and on the farm training of GAPS has been shown to strongly increase agricultural production by 2.6, 0.3, and 2.8 percent, respectively (Tambi, 2019). Unfortunately, not all farmers follow through on GAP training. Thus predicting who will and who will not adopt climate smart techniques is ever more important. This will enable policy makers to determine the most vulnerable to weather shocks and thus food insecurity.

The challenge in predicting which farmers will adopt techniques is there is no one reason for why a farmer may adopt a technique. Therefore, it is difficult to weight the myriad of inputs that go into a farmer's decision. Further, adoption happens with a lag, and is often not observable when using only one indicator.

Studies identify variables to predict which farmers will adopt climate smart practices, but have not been able to pinpoint a perfect solution due to the complexities of farmers' characteristics. (Baynes et al., 2011; Wonde et al., 2022). Additional findings show other than crop yield, the eager engagement in training and the application of acquired knowledge and skills are driven by the market incentives farmers anticipate for their produce as well. Therefore, regular and thorough supervision and evaluation are necessary to ensure the provided training meets the requirement, is delivered in a timely manner, and effectively brings about the desired impact (Wonde et al., 2022).

In this study, we aim to predict the adoption of agricultural practices. For example, we consider trainings on irrigation, seed intercropping, land preparation, input use, pest management, crop rotation to be climate smart agricultural practices of interest. Predicting who will adopt GAPs is important for reducing poverty, preventing climate change, and efficiently employing resources. We begin by building a model to predict the adoption of climate savings practices in Cambodia, specifically, the adoption of climate smart technologies in rice production. We extend this to predicting the adoption for barley farmers in Ethiopia to compare the performance of the models on a different country and crop. Utilizing both the Ethiopian data and Cambodian data, we perform cross-country analysis with a merged dataset to test generalizable performance.

We use survey data collected from the World Bank, as well as publically available data including Normalized Difference Vegetation Index (NDVI) and Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) to reflect the environmental conditions for the farmers. The data are from a 2022 endline survey collected 3 years after baseline. This data set includes farmer characteristics, geographic data, and a training impact variable. The Cambodian Amru data include 341 variables and 709 observations. Importantly, it includes demographic data to prevent unknowing discrimination on the basis of gender, age, or marital status through sources such as algorithmic

bias, human bias, selection bias, etc. Merging the latter data, we train several models to predict adoption out of sample: Ordinary Least Squares (OLS), Random Forest (RF), Ridge Regressions (RR), Neural Networks (NN), and Support Vector Machine (SVM) and evaluate the accuracy of our predictions on held out test data within country as well test data across country.

Our analysis reveals the efficacy of non linear models in predicting farmers' adoption of improved agricultural techniques in Cambodia and Ethiopia including Support Vector Machines (SVM), Random Forests (RF) and Neural Networks (NN). These models outperform linear models and exhibit the most accurate predictions. For single country datasets, SVM demonstrates superior prediction ability. For cross-country datasets, NN and RF show promising precision metrics, considering the limitations of the model. However, Ridge Regression, despite having a higher R-squared in Ethiopia, doesn't perform as well in terms of precision and accuracy. Additional findings underscore the importance of specific covariates, such as NDVI, latitude, longitude, age, and amount of land farmers use for practices in enhancing the models' predictive capacity.

# 2 Literature Review

This paper compares methods to identify factors that influence uptake of climate smart practices. To date, much of the literature identifies the most effective interventions and evaluates their impact on farmers' livelihoods (Jones et al., 2022; Silva et al., 2022; Stewart et al., 2015). However, there is a growing recognition that understanding the factors that influence uptake is equally important for resource allocation.

For instance, social networks influence in the uptake of a soil fertility management intervention among smallholder farmers in Kenya (Mponela et al., 2023). Previous studies have highlighted the complex interplay of factors that affect intervention uptake, including socio-economic context, cultural norms, and individual agency. (Baynes et al., 2011; Bekele et al., 2021; Bizikova et al., 2020; Frimpong-Manso et al., 2022). These findings from a handful of studies suggest that a nuanced understanding of the social and cultural context is needed to predict the effectiveness of crop farming interventions. To account for this nuance, we will use household level data, which captures context specific social factors augmented with publicly available satellite imagery that reflects temporal trends.

The development of reliable models to predict who will be the most impacted by GAP training can improve the livelihoods of smallholder farmers and maximize the overall effectiveness of treatment efforts. The scope of our model begins with Cambodia and its corresponding training interventions on rice cultivation such as irrigation, fertilizer, and pest management in an effort to improve food security. Even with increased economic growth, Cambodia's rates of malnutrition, anemia and nutrition-related deficiencies are high (Windus et al., 2022). Beyond Cambodia, in developing countries, smallholder farmers play an important role as they source up to 80% of food (IFADF, 2012). Subsequently, only inclusive rural transformation, such as interventions for improved GAP, can alleviate poverty in developing countries (IFADF, 2012).

We also contribute to a growing literature in economics that uses satellite data for both prediction and inferential questions (Allen et al., 2011; Nguyen et al., 2020; Wagner and Oppelt, 2020). In

this paper, we use satellite data to examine effects of agricultural training for a more refined and comprehensive evaluation. The estimation of crop yields with satellite data is fairly developed, however, only a few studies have estimated the lasting impacts of agricultural practices (Kubitza et al., 2020). Specifically, we use the Normalized Difference Vegetation Index (NDVI) as studies suggest NDVI is an adequate indicator for evaluating the impact of climate smart practices on plant health (Lebrini et al., 2020; Tamás et al., 2023). In our study, we can use the advantages of remote sensing data combined with sophisticated machine learning techniques to build robust and effective models.

A natural question arises concerning why, if integrating satellite data in predicting long-term adoption of climate smart practices is beneficial, it hasn't been widely implemented. We address this in a multi-faceted way. First, traditional methods, such as on-the-ground surveys and observational studies, satisfy researchers and policymakers to some extent. These methods provide a context-specific understanding of the situation and socio-economic factors influencing adoption (Ngowi et al., 2001; Yamoah and Kaba, 2022). Secondly, challenges related to the collection of reliable and accurate endline data and GPS coordinates hinder widespread implementation (Nakalembe, 2020). Lastly, complications arise when implementing machine learning models that utilize satellite data for prediction of long-term adoption. Researchers must meticulously optimize the parameters of these models for the best prediction, a process that demands computational resources and expertise (Gambella et al., 2021). Nevertheless, these barriers do not diminish the importance of integrating satellite data for prediction in this area. As improvements in remote sensing technology and computational capabilities continue, it becomes increasingly feasible to construct models that predict the adoption and effectiveness of climate smart practices. Such models could lead to a more efficient allocation of resources, improving the livelihoods of smallholder farmers and enhancing food security (IFADF, 2012).

This paper proceeds as follows: Section III describes the data used and potential biases affecting our results. Section IV explains our models and estimation procedure. Section V describes the results of our analysis and presents the best performing models. Section VI summarizes our findings and presents areas for future research.

# 3 Data Section

In 2017, the middle fifth quantile of rural households in Cambodia had an average disposable income per month of 332,000 riels per month as compared to 444,000 riels per month for their urban counterparts (excluding Phnom Penh, the capital of Cambodia) (NISMPD, 2017). Using the average exchange rate of 4,058.35 riels/USD in 2017 (exr, 2017), this meant Cambodians had an average disposable income of $81.8 or $109.4 for rural and urban households respectively. With almost 61 percent of Cambodians in rural areas, with 77 percent of those relying on agriculture, fisheries, and forestry, the Cambodian government is interested in improving agricultural training and the uptake of yield-enhancing and climate-smart practices. Several initiatives are in place to invest in farmers and to stabilize farm gate prices for rice producers in particular. One such initiative, ASPIRE, is a 7-year program from the Royal Government of Cambodia to foster innovation, resilience, and reduce poverty (asp, 2022). To reduce the impact of Covid-19, the World Bank also approved a $20 million project lasting from 2020-2024 to prevent, detect, and respond to the virus (Cam, 2020).

In addition to this, Germany and the World Food Program is providing cash assistance to aid Cambodian farmers affected by Covid-19 and climate shocks (WFP, 2021).

This study employs data provided by the International Finance Corporation sector of the World Bank, who partnered with two major Cambodian food companies with the goal of improving farmers' livelihoods by analyzing their use of Good Agricultural Practices (GAP). Mars Food, is the "owner of the world's largest rice brand Uncle Ben's, along with its local rice supplier Battambang Rice Investment Co., Ltd (BRICo)," (mar, 2019), which is one of the two companies in the Cambodian partnership. Amru Rice Cambodia Co., the other company in the partnership, is "the largest producer and exporter of organic rice from Cambodia" (amr, 2020). Baseline and endline data were collected to see the changes between the characteristics with the training interventions. Our study will be primarily focusing on the endline data to create a predictive model for adoption and determine which variables lead to a lasting impact of training interventions.

The IFC surveyed households in September 2019 for baseline data collection as well as in June 2022 for endline data collection. Most of the data is based on the household surveys, but the data on productivity and production costs comes from expert Focus Group Discussions (FGDs), which reflect the 2019 wet season. Initially, the IFC team conducted work cleaning the data in Stata to rename the variables and improve clarity. Our team continued to clean the dataset and merged the dependent variable of interest with the cleaned dataset, which is defined as a binary variable that answers the question: "Did you make any lasting changes to your farming practices using what you learned in training?" The specific covariates from the survey data are defined in Table 3. After modifying the survey data for the Amru dataset, there are 709 observations, which represent households and there are 341 covariates. We analyze the survey data provided in a Jupyter notebook using Python.

The data is divided into treatment and control groups, where the definitions of each group depend on the training provided and the specific intervention targeted. The objectives with the Cambodian Amru data are to support sustainable farming practices through (1) use of improved seeds, (2) better functioning cooperatives with focus on management of seed multiplication and facilitation of farmer access to improved quality seeds, (3) increased income from rice/reduced costs, (4) improved paddy quality, (5) better access to export markets, and (6) improved awareness and adoption of recommended practices related to gender. Therefore, our model allows the World Bank to have more predictive power and manage an efficient and effective use of resources to target specific characteristics of farmers and geographic areas where farmers will benefit most from lasting impact of the training.

In addition to Cambodian data, we use Ethiopian data on barley farmers for a single country analysis as well as for a cross-country analysis with Cambodian Mars data. The Ethiopian data contains similar covariates to the ones described in Table 3, but only has 165 independent variables with 1311 observations. Due to merging constraints, the Ethiopian dependent variable is slightly different, as we use a binary indicator from survey data, where farmers are asked: "Are you (or another household member) in need of training on improved agricultural practices in malt barley production?" Additionally, the dependent variable for Cambodian Mars is slightly different as it is also a binary indicator if farmers mastered at least 3/16 trainings. Although, it would be ideal to have perfectly standardized surveys, these variables serve as a proxy for long term uptake of GAP

adoption given the time constraints of the project.

To supplement the survey data, we incorporate publicly available datasets to enhance our predictive model. One such data source is CHIRPS provided by the University of California, Santa Barbara. CHIRPS is a high-resolution, quasi-global rainfall dataset that combines satellite imagery with in-situ station data to generate accurate precipitation estimates. By integrating the CHIRPS data into our model, we can better understand the influence of rainfall patterns and climatic conditions on farmers' adoption of the training interventions. This additional layer of information will not only improve the predictive power of our model but also help the World Bank and partner organizations to design targeted interventions that consider the potential impact of climate variability on farmers' decision-making processes. We only implement the CHIRPS data for the Cambodian Amru single country analysis due to availability of data.

Our study uses NDVI data in our models to make predictions about specific geographic areas to target when implementing treatment training for improved GAP. NDVI is "the quantitative index of greenness ranging from 0-1 where 0 represents minimal or no greenness and 1 represents maximum greenness" (Wasser and Holdgraf, 2021). NDVI is calculated from the visible and near-infrared light reflected by vegetation. Healthy vegetation absorbs most of the visible light that hits it, and reflects a large portion of the near-infrared light. Unhealthy or sparse vegetation reflects more visible light and less near-infrared light. Therefore, using the NDVI index with the latitude and longitudes from the household surveys, we can develop a more comprehensive model that will show visible effects of the treatment as well as a measure to compare to the self reported effects that the farmers report.

We face several caveats to using this survey data to predict adoption. Firstly, our samples may not be generalizable to other developing countries. This can be due to cultural, socioeconomic, or geographic factors that vary by country. Furthermore, selection bias may have skewed our results if certain groups were overrepresented or underrepresented in the sample. Social desirability bias may have also played a significant role as respondents may provide responses that they believe are socially acceptable or desirable rather than their true beliefs or behaviors, which could affect the adoption estimates. Lastly, recall bias could also impact the reliability of our data, as participants might not accurately remember past events or behaviors. They may forget, misremember, or selectively recall information, leading to inaccurate representations of their true behaviors or intentions. Selection bias, social desirability bias, and recall bias must all be addressed at the time of survey collection and are difficult to account for ex-post.

The implications of our policy recommendations may not apply to farmers who cultivate crops other than those studied, or to farmers operating in different international contexts where weather conditions and farming practices diverge. However, we gain promising results for generalizability in cross-country analysis. Furthermore, potential biases, such as selection and measurement biases, might exist in our data, limiting the broader applicability of our findings. If, for instance, our sample of Cambodian rice farmers does not reflect the wide spectrum of farmers' characteristics—including factors such as gender, caste, education, social and economic capital, farmland characteristics, access to the market, extension services, training, and major climate risks (Aryal et al., 2018), then our policy recommendations may not effectively address the diverse needs and circumstances of the entire population of rice farmers in Cambodia, thereby affecting their decision to adopt new technologies.

Nevertheless, a wide range of stakeholders stand to benefit from our research findings. These include policymakers shaping the agricultural landscape, investors in the agricultural sector, fellow researchers seeking to further refine these findings, the wider Cambodian public, and most importantly, the farmers whose livelihoods are directly impacted by these policy recommendations.

# 4 Problem and Estimation

It is difficult to predict which farmers will have a lasting uptake of climate smart practices because farmers are influenced by a multitude of factors and long-term data is limited. Since farming is dynamic, a slight change in latitude can change how a farmer would implement GAP. Additionally, inaccurate survey data, data biases, multicollinearity, heteroskedasticity, and computational limitations complicates our model. We address these concerns by comparing a range of models using our comprehensive survey data, NDVI, and CHIRPS as covariates to predict which farmers self report adoption of any GAP. The models we estimate include (1) Ordinary Least Squares (OLS), (2) Ridge Regression (RR), (3) Random Forest (RF), (4) Support Vector Machine (SVM), and (5) Neural Networks (NN).

Our outcomes of interest are famers' perceived effectiveness of training, and stated adoption of practices. Our covariate space consists of 341 covariates from demographic survey data NDVI, and rainfall data from CHIRPS.

The survey data includes responses of households about demographics, farming practices, costs, and production. Additionally, we are particularly interested in questions about training practices to understand if farmers attended the training and if so, what their goals with regard to the training. The training category covers questions about pre-harvest and post-harvest periods.

NDVI quantifies vegetation greenness and density, allowing us to predict adoption based on farmers' land health. Additionally, precipitation can impact adoption decisions and affects the cost of irrigation.

We train five distinctly different models on our data using k-fold cross validation with grid search. Overfitting is controlled for by fitting the model on several folds of the training data and predicting on held-out test data. For all machine learning models we tune the hyperparameters to prevent overfitting.

(1) OLS

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_n X_{ni} + \epsilon_i \tag{1}$$

where $Y_i$ is the dependent variable for the $i$th observation. This means $Y_i$ is the comprehensive lasting impact of adopting various training on $i$th farmer's agricultural practice. The subscript $i$ describes the ith person's characteristics. This is the respondent of the survey, who answers for their household. $\alpha$ is the constant term, $X_{1i}, X_{2i} \ldots, X_{ni}$ are the independent variables for

the $i$th observation. $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients to be estimated, and $\epsilon_i$ is the error term for the $i$th observation which is assumed to be normally distributed with mean 0 and constant variance.

For estimation, the following assumptions are imposed: (1) Linear relationship: We assume that the relationship between the dependent variable and independent variables is linear; (2) Independence: Observations are independent of each other; (3) Homoskedasticity: The error term has constant variance across all levels of the independent variables; (4) Normality: The error term is normally distributed. This model yields poor results as it yields an $R^2$ value of $-8.132$, which is unsurprising as the assumptions above do not hold for our model.

(2) RR

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p \tag{2}$$

Where the coefficients are chosen to minimize this objective function are:

$$\sum_{i=1}^{n} (Y_i - \hat{Y}i)^2 + \lambda \sum j = 1^p \beta_j^2 \tag{3}$$

where $\lambda$ is the regularization parameter

For the estimation of RR, we impose the same assumptions as OLS except that Ridge does not assume the absence of multicollinearity.

(3) RF

First, in estimating the RF model, we randomly select a subset of the training data and a subset of the predictor variables. Then, we grow a decision tree on the selected subset of the data and predictor variables using Gini impurity as splitting criterion. These steps are repeated multiple times to create a forest of decision trees. For a new observation, we predict its outcome by averaging the predictions of all decision trees in the forest. Therefore, after splitting the data into training and test sets, we fitted the model to make a prediction on the test data and plotted to visualize it. Our model uses 500 trees and tries 165 variables at each split. The only assumption in the RF model is that sampling is representative.

(4) SVM

In the SVM model, we split the data between training and test sets. Another paper implements SVM and describes it as the following: "In a regression SVM model, you have to estimate the functional dependence of the dependent variable y on a set of independent variables x" (Noori et al., 2011).

Given a set of training data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, where $x_i$ is a feature vector and $y_i$ is the class label (1 or -1), the SVM algorithm tries to find the hyperplane $w^T x + b = 0$ that maximizes the margin between the positive and negative samples, subject to the constraint that all samples are classified correctly.

The margin is given by $2/||w||$, where $||w||$ is the Euclidean norm of the weight vector $w$. The objective function to be optimized is therefore:

$$\min_{w,b} \frac{1}{2}||w||^2 \tag{4}$$

subject to $y_i(w^T x_i + b) \geq 1$ for $i = 1, 2, \ldots, n$ where the inequality constraint ensures that all samples are classified correctly (Noori et al., 2011).
In addition to the hyperplane, we also use a kernel to allow the SVM algorithm to solve non-linear problems by implicitly transforming the data to a higher-dimensional space without actually computing the transformation explicitly.

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{5}$$

subject to $\sum_{i=1}^{n} \alpha_i y_i = 0$ and $0 \leq \alpha_i \leq C$ for all $i$, where $K(x_i, x_j)$ is the kernel function and $\alpha_i$ are the Lagrange multipliers.

The kernel function $K(x_i, x_j)$ maps the data to a higher dimensional feature space where it might be linearly separable. The most common kernel functions are:

Linear: $K(x_i, x_j) = x_i^T x_j$ Polynomial: $K(x_i, x_j) = (x_i^T x_j + c)^d$ Radial basis function: $K(x_i, x_j) = e^{-\gamma||x_i - x_j||^2}$ The decision function for a new sample $x$ is given by:

$$f(x) = \text{sgn}\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b\right) \tag{6}$$

This function allows us to make predictions for new data points after the model has been trained. Here, the term sgn represents the sign function. For our optimization, we choose the linear kernel for its scalability, robustness to noise, and interactions between features.

(5) NN

The human brain inspires Neural Networks (NN), designed to learn complex patterns in data by iteratively adjusting the connections (weights) between neurons (nodes) to minimize the error between predicted and actual outputs. Neural networks consist of an input layer, one or more hidden layers, and an output layer, with each layer containing nodes connected to adjacent layers through weighted connections. The fundamental building block is the artificial neuron, represented as

$$y = f(\sum(w_i \times x_i) + b) \tag{7}$$

where $y$ is the output,$f$ is the activation function, $w_i$ are the weights, $x_i$ are the inputs, and $b$ is the bias term.

To predict which farmers will continue to implement training programs in the long term, a neural network can be used as a binary classifier, with input features such as farm size, previous training programs attended, and years of experience. The output layer would have one neuron with a sigmoid activation function, providing a probability value. Training involves backpropagation, minimizing the error using the mean squared error (MSE) as the loss function:

$$L(y, \hat{y}) = \frac{1}{n} \sum (y_i - \hat{y}_i)^2 \tag{8}$$

After training, the model can predict the likelihood of farmers continuing to implement training programs, and its performance can be evaluated using accuracy, precision, recall, and F1 score. This model allows for the highest degree of nonlinearity but is most prone to overfitting.

Our models' accuracy will be assessed using k-fold cross-validation. We compare with the mean-squared error (MSE). The MSE measures the difference between the actual and predicted values. A lower MSE indicates better model performance.

Our 5-fold cross-validation, assesses the performance of a model on new, unseen data. By running multiple folds, we can balance the bias-variance tradeoff in prediction and control for overfitting.

To perform K-fold cross validation, we first divide the dataset into 'k' equal parts or "folds". The model is then trained on 'k-1' folds and tested on the remaining held out fold. This process is repeated 'k' times, with each fold being used as a test set only once. By averaging the MSE across all 'k' iterations, we obtain a more accurate estimation of the model's performance, as well as a means by which we can select the optimal hyperameters.

To the best of our knowledge there are few papers that aim to prediction the adoption of agricultural practices. Llewellyn and Brown (2020) stands out in the relatively sparse research landscape of predicting adoption of agricultural practices by farmers, particularly in the context of smallholder agriculture in developing countries. Their work underscores the distinct factors influencing these farmers, such as the heterogeneity in their resources, capabilities, and priorities, cultural influences, and access to agricultural extension services (Llewellyn and Brown, 2020).

In addition to these factors, a paper on smallholder farmers in Malawi finds neighborhood effects and incentives were strong predictors in farming practice adoption (Bell et al., 2018). Their research also emphasizes the potential of subsistence-oriented farming objectives, short-term returns, upfront costs, and sociocultural characteristics to influence the exposure to innovations and their perceived relative advantage. The study recognizes the challenge of predicting adoption due to unique features of smallholder agriculture that may slow down diffusion and decrease the relative advantage of new practices. We believe that our comprehensive survey data paired with NDVI and CHIRPS does a sufficient job of combining both quantitative and qualitative factors unique to smallholder farmers and will allow us to achieve an accurate and nuanced prediction.

# 5 Results

The results highlight how Support Vector Machines (SVM) and Neural Networks (NN) effectively predict farmers' adoption of improved agricultural techniques in Cambodia and Ethiopia based on the metrics of precision and accuracy. Detailed discussions follow, exploring each model's individual performance, the significant features influencing prediction accuracy, and the outcome of cross-validation when swapping training and testing sets between the two countries.

We begin with a simple linear model for both countries, with metrics shown in column 1 of table 1, and find it does not perform well out of sample. The R-squared is negative indicating that a model of just the mean outcome would better predict adoption. This is not surprising given the high collinearity between our covariates and the tendency for linear models without penalization to overfit the training sample, and, thus, perform poorly on out of sample data. We progress to models that account for non-linearities in the data.

Panel A in table 1 show the training results for different models in Cambodia: Ridge Regression (RR), Support Vector Machines (SVM), Random Forest (RF), and Neural Networks (NN). These address the concerns that linear models present, which include assume linear relationships between the features and outcomes. SVM emerges as the most efficient model, recording the highest R-squared (0.5614), accuracy (0.9155), and precision (0.9346). This implies SVM's superiority in predicting new rice farming technique adoption in Cambodia. Conversely, RR records a remarkably low R-squared (-146.2062) and high MSE (27.7048), indicating probable overfitting and emphasizing the importance of regularization techniques in predictive models.

For the Ethiopian dataset, with metrics displayed in panel B of table 1, the Random Forest model stands out, achieving a high R-squared value of 0.9734 along with high accuracy (0.9949) and precision (0.9950). However, while RR yields a higher R-squared (0.7516) compared to other models except for RF, it performs poorly in terms of accuracy and precision. The discrepancy may arise from the specifications of the models and the purpose of the metrics. Accuracy will reflect the correctness of class predictions, while R-squared measures the explained variance in the predicted outcome. The SVM model maximizes the margin between classes in the dataset without explicitly aiming to minimize variance in the response variable or maximize R-squared while Neural Networks optimize for minimizing error in predictions rather than explaining variance, which is the aim of R-squared. Ridge Regression differs from these two models as it penalizes large coefficients and addresses multicollinearity through regularization, which affects the R-squared value, not precision or accuracy of accurately predicting adoption or non-adoption.

To enhance the comprehension of our models' predictive capabilities, we visualize the prediction results from our NN model for both countries on Google map. We divide the results into three categories based on each farmer's response and prediction accuracy. "Positive" stands the respondent answers yes to the question of whether adopt new agricultural techniques and make long-term changes. "Negative" stands the respondent answers no. Figure 1 uses Cambodia data with rice and figure 2 uses Ethiopia data with barley. Both visualization maps demonstrate again how effective NN model is.

### 5.1 Robustness Check

One concern with predictive models using small survey datasets is their ability to work across country contexts. In this section, we fit a similar predictive model to a new dataset of barley farmers in Ethiopia, which predicts which farmers adopt GAP long-term. We train the model using one country's dataset and test it on the other country's data out of sample. Table 2 summarizes our results. Although our performance metrics are poorer than those in Table 1, we achieve promising results considering we test the model with data from a different continent with different crop and train with only 45 covariates. We conclude the precision metrics in the RF and NN model in panel B of table 2 show the best-performance with the cross-country data. Panel B, with Ethiopia as a training set, likely performs better than Panel A due to the imbalance in number of observations in the datasets, 1311 for Ethiopia and 856 for Cambodia (Mars). With precision measures of over 76% in both models, one would be able to reliably predict true positives, the farmers who uptake long-term adoption of GAP. These precision metrics would likely increase with an increase in the number of covariates and the sample size. Therefore, project leaders could administer trainings to farmers who would truly benefit from it, likely leading to an increase in cost-effectiveness, farmers' food security, and global sustainability.

### 5.2 Additional Findings

In addition to applying our prediction framework to an alternate country, we explore which covariates most contribute to the predictive capacity our models. We implement the technique of feature importance from RF, showing how much each feature contributes to reducing the impurity in the dataset when constructing the decision trees within the ensemble. We use RF for the feature importance due to its high performance metrics in all models we construct. Due to the large number of covariates, all features show relatively low importance in single country models shown with Cambodia in Figure 3 and Ethiopia in Figure 4. However, with the limited number of covariates in the cross-country dataset, we are able to conclude the top covariates for explaining adoption were longitude, latitude, NDVI, amount of land used, and age of the respondent, shown in Figure 5.

We also examine feature importance through Shapley values, which measure the average contribution of covariates across all possible combinations of features in a prediction. Due to computational intensity and time constraints, we only calculate Shapley values for the cross-country dataset, since we include 44 covariates in the model, displayed in Figure 6. We conclude a similar finding as Random Forest feature importance where the variables challenge of quality seeds, Longitude, challenge of weather, Latitude, and amount of land used have the highest Shapley values.

## 6 Conclusion

This research predicts the adoption of GAP by farmers in two developing countries, specifically Cambodia and Ethiopia. We utilize survey data supplemented by two additional climate indices: CHIRPS and NDVI data to predict the adoption of GAP. The survey data, collected over three years post-training, provides detailed insights into farmers' socioeconomic backgrounds, their resources, and their attitudes towards new farming practices. Through a combination of traditional statistical models and machine learning methods, including OLS, RR, RF, NN, and SVM, we train several

models to predict which farmers self report adoption of any GAP. We first hypertune our models on training data, and test each model on held-out test data for each country. We also study whether a trained model on one country's data can perform well in predicting adoption out-of-sample for another country.

We begin by fitting each of our models, where SVM and NN perform the best. We achieve over 84% predictive power with our SVM and NN models in terms of precision and accuracy. We obtained the highest values of precision and accuracy in our NN model at 96.94% and 92.25% respectively. The SVM model, due to its ability to handle multiple continuous and categorical variables, also exhibits high accuracy. We conduct a secondary analysis on Ethiopian data, which supports our primary findings - namely, that SVM and NN are the best performing. This further empowers our approach and provides evidence for our model to be generalized to other crops and countries.

We conduct an additional exercise to test whether our models not only perform well out of sample for a given country, but also if a trained model for Cambodia can predict the adoption of GAP in Ethiopia and vice versa. When we train the model using Ethiopian data, we achieve a R-squared of -8.76 and accuracy of 0.12 for NN and when we train it with Cambodian data, we achieve a R-squared of -1.81 and accuracy of 0.44 for NN. However, we obtain promising precision metrics for the RF and NN models, which would likely increase with more covariates and more similar data.

Overall, our findings suggest that a combination of comprehensive survey data and environmental data, processed through NN or SVM models, can accurately predict the adoption of good agricultural practices by farmers in developing countries.

There are several caveats to our findings. First, we are limited to data sets that had latitude and longitude data, which are needed for precipitation and NDVI values. Within these data, we drop or recategorize missing observations, slightly decreasing our sample size. While our model performed well on Cambodian and Ethiopian data, its generalizability across country was less favorable, but still promising. One of the difficulties in performing robustness checks with our models is the lack of standardized survey questions. Not all surveys ask the same question, and therefore, running the same model with a different survey limited the number of variables we could use to test a model on an out-of-sample data set for another country.

Future work can include additional machine learning models, including ensemble methods, which could be beneficial to predictive performance. Our study focuses on SVM and NN, but other methods such as gradient boosting or models that combine predictions may prove more accurate. For future work, one could perform more data cleaning to merge the surveys from multiple countries into a set of standardized covariates in order to improve the number of covariates used for cross-country predictions. Our study also only focuses on adoption, where other measures may be important for improving the livelihoods of farmers, for example, crop yields, income, wealth, or reducing environmental impact. In this way, expected values for training programs can be calculated and policy decisions can be weighed.

In conclusion, this study predicts the adoption of good agricultural practices using machine learning techniques. The model we built can be used for any other country or other products with appropriate data, which can be implemented for greater policy implementations, not only Cambodian or

Ethiopian farmers. We hope our work will inspire future research in this area and contribute to the ongoing efforts to improve the lives of farmers and mitigate the effects of climate change on agriculture.

# References

(2017). Us dollar (usd) to cambodian riel (khr) exchange rate history for 2017. Website. Accessed May 29, 2023.

(2019). A 'win-win' situation for sustainable rice. Website. Accessed May 30, 2023.

(2020). Amru rice. Website. Accessed May 30, 2023.

(2020). Cambodia covid-19 emergency response project. Website. Accessed May 30, 2023.

(2021). Germany and world food programme provide cash assistance to support recovery of vulnerable cambodian. Website. Accessed May 30, 2023.

(2022). Aspirekh. Website. Accessed May 29, 2023.

Allen, R., A. Irmak, R. Trezza, J. M. Hendrickx, W. Bastiaanssen, and J. Kjaersgaard (2011). Satellite-based et estimation in agriculture using sebal and metric. *Hydrological Processes 25*(26), 4011–4027.

Aryal, J. P., M. L. Jat, T. B. Sapkota, A. Khatri-Chhetri, M. Kassie, S. Maharjan, et al. (2018). Adoption of multiple climate-smart agricultural practices in the gangetic plains of bihar, india. *International Journal of Climate Change Strategies and Management*.

Baynes, J., J. Herbohn, I. Russell, and C. Smith (2011). Bringing agroforestry technology to farmers in the philippines: Identifying constraints to the success of extension activities using systems modelling. *Small-scale Forestry 10*, 357–376.

Bekele, R. D., A. Mirzabaev, and D. Mekonnen (2021). Adoption of multiple sustainable land management practices among irrigator rural farm households of ethiopia. *Land Degradation & Development 32*(17), 5052–5068.

Bell, A. R., J. Zavaleta Cheek, F. Mataya, and P. S. Ward (2018). Do as they did: Peer effects explain adoption of conservation agriculture in malawi. *Water 10*(1), 51.

Bizikova, L., E. Nkonya, M. Minah, M. Hanisch, R. M. R. Turaga, C. I. Speranza, M. Karthikeyan, L. Tang, K. Ghezzi-Kopel, J. Kelly, et al. (2020). A scoping review of the contributions of farmers' organizations to smallholder agriculture. *Nature Food 1*(10), 620–630.

Frimpong-Manso, J., E. K. Tham-Agyekum, D. C. Aidoo, D. Boansi, E. O. Jones, and J.-E. A. Bakang (2022). Cooperative membership status and adoption of good agronomic practices: Empirical evidence from cocoa farmers in atwima mponua district, ghana. *The Bangladesh Journal of Agricultural Economics 43*(1), 1–17.

Gambella, C., B. Ghaddar, and J. Naoum-Sawaya (2021). Optimization problems for machine learning: A survey. *European Journal of Operational Research 290*(3), 807–828.

IFADF (2012). Annual report 2012.

Jones, M., F. Kondylis, J. Loeser, and J. Magruder (2022). Factor market failures and the adoption of irrigation in rwanda. *American Economic Review 112*(7), 2316–52.

Kansanga, M. M., I. Luginaah, R. B. Kerr, L. Dakishoni, and E. Lupafya (2021). Determinants of smallholder farmers' adoption of short-term and long-term sustainable land management practices. *Renewable agriculture and food systems 36*(3), 265–277.

Ki-moon, B., K. Georgieva, and B. Gates (2019). Adapt now: a global call for leadership on climate resilience. *Global Commission on Adaptation*, 90.

Kubitza, C., V. V. Krishna, U. Schulthess, and M. Jain (2020). Estimating adoption and impacts of agricultural management practices in developing countries using satellite data. a scoping review. *Agronomy for Sustainable Development 40*, 1–21.

Lebrini, Y., A. Boudhar, A. Htitiou, R. Hadria, H. Lionboui, L. Bounoua, and T. Benabdelouahab (2020). Remote monitoring of agricultural systems using ndvi time series and machine learning methods: a tool for an adaptive agricultural policy. *Arabian Journal of Geosciences 13*(16), 796.

Llewellyn, R. S. and B. Brown (2020). Predicting adoption of innovations by farmers: What is different in smallholder agriculture? *Applied Economic Perspectives and Policy 42*(1), 100–112.

Moser, C. M. and C. B. Barrett (2003). The disappointing adoption dynamics of a yield-increasing, low external-input technology: the case of sri in madagascar. *Agricultural systems 76*(3), 1085–1100.

Mponela, P., J. Manda, M. Kinyua, and J. Kihara (2023). Participatory action research, social networks, and gender influence soil fertility management in tanzania. *Systemic Practice and Action Research 36*(1), 141–163.

Nakalembe, C. (2020). Urgent and critical need for sub-saharan african countries to invest in earth observation-based agricultural early warning and monitoring systems. *Environmental Research Letters 15*(12), 121002.

Ngowi, A. V., D. N. Maeda, C. Wesseling, T. J. Partanen, M. P. Sanga, and G. Mbise (2001). Pesticide-handling practices in agriculture in tanzania: Observational data from 27 coffee and cotton farms. *International journal of occupational and environmental health 7*(4), 326–332.

Nguyen, T. T., T. D. Hoang, M. T. Pham, T. T. Vu, T. H. Nguyen, Q.-T. Huynh, and J. Jo (2020). Monitoring agriculture areas with satellite images and deep learning. *Applied Soft Computing 95*, 106565.

NISMPD (2017). Cambodia socio-economic survey 2017. Website.

Noori, R., A. Karbassi, A. Moghaddamnia, D. Han, M. Zokaei-Ashtiani, A. Farokhnia, and M. G. Gousheh (2011). Assessment of input variables determination on the svm model performance using pca, gamma test, and forward selection techniques for monthly stream flow prediction. *Journal of hydrology 401*(3-4), 177–189.

Silva, J. V., V. O. Pede, A. M. Radanielson, W. Kodama, A. Duarte, A. H. de Guia, A. J. B. Malabayabas, A. B. Pustika, N. Argosubekti, D. Vithoonjit, et al. (2022). Revisiting yield gaps and the scope for sustainable intensification for irrigated lowland rice in southeast asia. *Agricultural Systems 198*, 103383.

Stewart, R., L. Langer, N. R. Da Silva, E. Muchiri, H. Zaranyika, Y. Erasmus, N. Randall, S. Rafferty, M. Korth, N. Madinga, et al. (2015). The effects of training, innovation and new technology on african smallholder farmers' economic outcomes and food security: a systematic review. *Campbell Systematic Reviews 11*(1), 1–224.

Tamás, A., E. Kovács, É. Horváth, C. Juhász, L. Radócz, T. Rátonyi, and P. Ragán (2023). Assessment of ndvi dynamics of maize (zea mays l.) and its relation to grain yield in a polyfactorial experiment based on remote sensing. *Agriculture 13*(3), 689.

Tambi, M. D. (2019). Agricultural training and its impact on food crop production in cameroon. *Journal of Socioeconomics and Development 2*(1), 1–11.

Wagner, M. P. and N. Oppelt (2020). Extracting agricultural fields from remote sensing imagery using graph-based growing contours. *Remote sensing 12*(7), 1205.

Wasser, L. and C. Holdgraf (2021). Lesson 1. calculate vegetation indices in python.

Windus, J. L., K. Duncanson, T. L. Burrows, C. E. Collins, and M. E. Rollo (2022). Review of dietary assessment studies conducted among khmer populations living in cambodia. *Journal of Human Nutrition and Dietetics 35*(5), 901–918.

Wonde, K. M., A. S. Tsehay, and S. E. Lemma (2022). Training at farmers training centers and its impact on crop productivity and households' income in ethiopia: A propensity score matching (psm) analysis. *Heliyon 8*(7), e09837.

Yamoah, F. A. and J. S. Kaba (2022). Integrating climate-smart agri-innovative technology adoption and agribusiness management skills to improve the livelihoods of smallholder female cocoa farmers in ghana. *Climate and Development*, 1–7.

Yang, Q., Y. Zhu, and F. Wang (2021). Exploring mediating factors between agricultural training and farmers' adoption of drip fertigation system: Evidence from banana farmers in china. *Water 13*(10), 1364.

# 7 Tables

Table 1: Out of Sample Prediction on Test Data in Cambodia and and Ethiopia

| | *Panel A - OOS Prediction for Rice Farmers in Cambodia* | | | | |
|---|---|---|---|---|---|
| | OLS | Ridge | Random Forest | Neural Network | SVM |
| R-squared | -0.7976 | -146.2062 | 0.3990 | 0.4517 | 0.5614 |
| Accuracy | N/A | N/A | 0.8803 | 0.8944 | 0.9155 |
| Precision | N/A | N/A | 0.8780 | 0.9327 | 0.9346 |
| MSE | 0.3306 | 27.7048 | 0.1197 | 0.1056 | 0.0845 |
| | *Panel B - OOS Prediction for Barley Farmers in Ethiopia* | | | | |
| | OLS | Ridge | Random Forest | Neural Network | SVM |
| R-squared | 0.3994 | 0.7516 | 0.3603 | 0.6349 | 0.9734 |
| Accuracy | N/A | N/A | 0.8680 | 0.9315 | 0.9949 |
| Precision | N/A | N/A | 0.8649 | 0.9286 | 0.9950 |
| MSE | 0.2960 | 0.0503 | 0.1320 | 0.0761 | 0.0051 |

For rice farmers in Cambodia, SVM had the best accuracy, precision, and MSE. For barley farmers in Ethiopia, RF has the best accuracy, precision, and MSE.

Table 2: Out of Sample GAP Adoption Across Country

| | *Panel A: Cambodia as Training Set, Ethiopia as Test Set* | | | | |
|---|---|---|---|---|---|
| | OLS | Ridge | Random Forest | Neural Network | SVM |
| R-squared | -21.8319 | -25.0754 | -2.5882 | -1.8088 | -2.5919 |
| Accuracy | N/A | N/A | 0.2777 | 0.4355 | 0.2768 |
| Precision | N/A | N/A | 0.4580 | 0.5798 | 0.0767 |
| MSE | 4.6255 | 6.5565 | 0.7269 | 0.5690 | 0.7277 |
| | *Panel B: Ethiopia as Training Set, Cambodia as Test Set* | | | | |
| | OLS | Ridge | Random Forest | Neural Network | SVM |
| R-squared | -2.3680 | -2055775840106351.2 | -8.4366 | -8.7594 | -8.5111 |
| Accuracy | N/A | N/A | 0.1121 | 0.1238 | 0.1051 |
| Precision | N/A | N/A | 0.7940 | 0.7626 | 0.0111 |
| MSE | 2.2280 | 516910261646707.4 | 0.8879 | 0.9182 | 0.8949 |

Merged Dataset with Cambodia MARS Data & Ethiopian Heineken Data
With Cambodia as the training set and Ethiopia as the test set, the NN has the best accuracy, precision, and MSE. When Ethiopia is the training set and Cambodia is the test set, the NN has the highest accuracy, but a slightly lower precision and MSE than the RF.

# 8 Appendix

Table 3: Survey Instrument and Summary Statistics

| Variables | Description | Mean (std) |
|---|---|---|
| Demographics/Other | | |
| age_resp | Age of respondent | 49.7825 |
| decision_maker | 1 if respondent is the decision maker for rice farming; 2 otherwise | 1.0099 (0.099) |
| treat | 1 if treatment farmer; 2 if control farmer | 1.4294 (0.4953) |
| district | 1. Sangkae; 2. Moung Ruessei; 3. Batheay; 4. Prey Chhor; 5. Choeung Prey; 6. Baray; 7. Prasat Ballangk; 8. Kampong Svay; 9. Stoung; 10. Santuk; 11. Prasat Sambour | 4.9294 (3.2266) |
| test_number | What is the missing number in this data series: 1,2,3,4,5,6,_, 8,9,10? 1 if correct answer; 2 if not correct answer | 1.065 (0.2467) |
| test_read | Please read out the following numbers: 54, 99, 208 | 1.065 (0.2467) |
| province | 1. Battambang; 2. Kampong Cham; 3. Kampong Thom | 2.0551 (0.795) |
| latitude | Latitude of farming area | 11.745 (3.0864) |
| literacy | 1 if respondent can read or write Khmer; 2 if respondent cannot | 1.1144 (0.3185) |
| longitude | Longitude of farming area | 97.839 (25.5184) |
| marital_status | 1 if never married; 2 if married; 3 if widower/widow; 4 if divorced; 5 if separated; 6 if married but unknown status | 2.0311 (0.322) |
| member | 1 if member of cooperative or farmer group; 2 if not a member; 3 if doesn't know | 1.4068 (0.6458) |
| gender | 1 if male; 2 if female; 3 if others | 1.5932 (0.4916) |
| irrigation | 1 if respondent does not use irrigation and rain fed only; 2 if respondent does | 1.3093 (0.4625) |
| 1. Gender Mapping Activities | In your household, who does the following activities in the rice field and at home? 1 if men only; 2 if mostly men; 3 if men and women equally; 4 if mostly women; 5 if women only; 6 if not applicable | |
| activity_buyinginputs | Buying agrochemicals and seeds | 2.5141 (1.1388) |
| activity_children | Look after children | 4.0777 (0.8713) |
| activity_fertilizing | Fertilizing | 2.2839 (1.1976) |
| activity_harvesting | Harvesting | 3.3475 (1.7666) |
| activity_housework | Housework | 3.9025 (0.923) |
| activity_irrigation | Water management/irrigation | 2.2571 (1.2298) |
| activity_landprep | Land preparation | 2.6088 (1.7826) |

| | | |
|---|---|---|
| activity_packing | Packing rice | 3.2726<br>(1.636) |
| activity_planting | Planting | 2.4082<br>(1.2017) |
| activity_selling | Selling rice | 3.1186<br>(1.1042) |
| activity_spraychem | Spraying chemicals | 2.2472<br>(1.5678) |
| activity_threshing | Threshing | 3.7331<br>(1.9888) |
| activity_transporting | Transporting rice from farm | 3.0169<br>(1.8561) |
| activity_weeding | Weeding | 2.4181<br>(1.3603) |
| 2. Challenges | What are the biggest challenges<br>you face when farming rice?<br>The respondent may select up to 3 options.<br>(1 if challenge; 0 otherwise) | |
| challenge_accessmarket | Access to market | 0.0254<br>(0.1575) |
| challenge_affordchem | Affordable agrochemicals | 0.0226<br>(0.1487) |
| challenge_affordloan | Affordable loans | 0.0014<br>(0.0376) |
| challenge_covid | Impacts of COVID-19 | 0.0042<br>(0.065) |
| challenge_disease | Crop disease | 0.2302<br>(0.4213) |
| challenge_efficacychem | Efficacy of agrochemicals | 0.0184<br>(0.1344) |
| challenge_farmadvise | Ability to get advice about farming | 0.0042<br>(0.065) |
| challenge_farmingskill | Lack of farming skills | 0.0918<br>(0.289) |
| challenge_getchem | Knowing where to buy agrochemicals | 0.0311<br>(0.1736) |
| challenge_labor | Not enough labor | 0.048<br>(0.214) |
| challenge_laborcost | Cost of labor | 0.0593<br>(0.2364) |
| challenge_other | Others | 0.113<br>(0.3168) |
| challenge_pest | Pest management | 0.2655<br>(0.4419) |
| challenge_postharv | Lack of post-harvest facilities (drying, milling) | 0.041<br>(0.1983) |
| challenge_priceseeds | Price of rice seeds | 0.0946<br>(0.2929) |
| challenge_qualityseeds | Quality of seeds | 0.0226<br>(0.1487) |
| challenge_unpredprice | Unpredictable market prices | 0.4562<br>(0.4984) |
| challenge_water | Lack of water for irrigation | 0.4463<br>(0.4975) |

| | | |
|---|---|---|
| challenge_weather | Weather conditions | 0.4901<br>(0.5003) |
| 3. Children | How many school aged children live in your household? | |
| child_1217 | Number of children age 12-17 | 0.5282<br>(0.7748) |
| child_611 | Number of children age 6-11 | 0.363<br>(0.7796) |
| child_school | Are there any school aged children who go to school? | 0.5282<br>(0.5324) |
| child_school_1217 | Number of children age 12-17 who attend school | 0.4534<br>(0.6743) |
| child_school_611 | Number of children age 6-11 who attend school | 0.3277<br>(0.5946) |
| 4. Costs | What production costs per 1ha of land did you incur in the 2021 wet season for the following? In Riel | |
| costha_equipment | Equipment hired (incl fuel) and used by this farm; possible range: 0-400,000 | 66476.6949<br>(100623.8768) |
| costha_fertilizer | Fertilizer; possible range: 0-1,000,000 | 378966.1325<br>(223129.1397) |
| costha_harvesting | Harvesting; possible range: 0-500,000 | 312129.2881<br>(84996.0391) |
| costha_irrigation | Irrigation; possible range: 0-200,000 | 2840.5191<br>(18854.3313) |
| costha_labour | Labor; possible range: 0-200,000 | 34913.1568<br>(49573.2547) |
| costha_landprep | Land preparation; possible range: 0-400,000 | 229838.6299<br>(92545.1312) |
| costha_other | Others | 423.7288<br>(11274.6904) |
| costha_othercosts | Others | 20834.7458<br>(99475.9797) |
| costha_pesticides | Pesticides; possible range: 0-400,000 | 66227.4082<br>(78458.1594) |
| costha_seeds | Seeds; possible range: 100,000-600,000 | 214129.2486<br>(119295.1862) |
| 5. Impact of COVID-19 | If the respondent answers yes to the COVID question, then surveyors will ask about specific impacts. | |
| COVID | Has the COVID-19 pandemic negatively affected your business? 1. Yes 2. No | 1.2712<br>(0.4449) |
| COVID_negeff_difficult_to_acce | 1 if it was more difficult to access inputs; 0 otherwise | 0.1582<br>(0.3652) |
| COVID_negeff_difficult_to_sell | 1 if it was more difficult to sell crops; 0 otherwise | 0.298<br>(0.4577) |
| COVID_negeff_I_worked_fewer_ho | 1 if the respondent worked fewer hours; 0 otherwise | 0.2669<br>(0.4427) |
| COVID_negeff_Other__SPECIFY_ | Other | 0.0367<br>(0.1882) |
| COVID_negeff_The_price_of_crop | 1 if the price of crops decreased; 0 otherwise | 0.5198<br>(0.5) |
| COVID_negeff_The_price_of_inpu | 1 if the price of inputs increased; 0 otherwise | 0.3814<br>(0.4861) |

| | | |
|---|---|---|
| 6. Rice Income Spending | How was additional income from rice in 2021 spent? | |
| addincome_rent_or_buy | 1 if rent or buy more land; 0 otherwise | 0.0424 (0.2016) |
| addincome_farm_equip | 1 if farming equipment; 0 otherwise | 0.0819 (0.2744) |
| addincome_buy_farm_inputs | 1 if buying farming inputs; 0 otherwise | 0.3517 (0.4778) |
| addincome_pay_debt | 1 if paying debt; 0 otherwise | 0.8008 (0.3996) |
| addincome_non_farm | 1 if doing non-farming business activities; 0 otherwise | 0.0664 (0.2491) |
| addincome_educ | 1 if children's education; 0 otherwise | 0.1836 (0.3874) |
| addincome_general | 1 if general household expenditures; 0 otherwise | 0.8121 (0.3909) |
| addincome_health | 1 if health care; 0 otherwise | 0.2119 (0.4089) |
| addincome_saving | 1 if savings; 0 otherwise | 0.0749 (0.2633) |
| 7. Ownership/use of farming equipment | Where did you get your farming equipment, such as tractor, plow, harvest machine...? | |
| equip_provider_I_borrow_i | 1 if respondent borrows it for free | 0.0099 (0.099) |
| equip_provider_I_own_it_a | 1 if respondent owns it and also shares it | 0.1102 (0.3133) |
| equip_provider_I_own_it_b | 1 if respondent owns it but doesn't share it | 0.137 (0.3441) |
| equip_provider_I_rent_it | 1 if respondent rents it | 0.9011 (0.2987) |
| equipment_available | Was farming equipment available when you need it? 1 if always; 2 if most of the time; 3 if sometimes | 1.5749 (0.8334) |
| 8. Fertilizer | Where did you get your fertilizers from? Select all that apply | |
| fert_provider_Cooperative | 1 if cooperative; 0 otherwise | 0.2472 (0.4317) |
| fert_provider_Directly_f | 1 if directly from rice company or processor; 0 otherwise | 0.0141 (0.1181) |
| fert_provider_Local_shop | 1 if local shop; 0 otherwise | 0.8588 (0.3485) |
| fert_provider_Other__SPE | Other | 0.0071 (0.0838) |
| fertilizer_chemical | 1 if farmer used chemical fertilizer; 0 otherwise | 1.0311 (0.1736) |
| fertilizer_organic | 1 if farmer used organic fertilizer; 0 otherwise | 1.2429 (0.4292) |

| | | |
|---|---|---|
| fertilizer_organic_how | Conditions: (1) Farmer could use it in non-flooded fields in composted or de-composted state; (2) Sufficient time for de-composition prior to flooding; (3) Available locally used fertilizer if (1) and (2) are met, but not (3); 2 if farmer did not use as fertilizer because one or more conditions could not be met; 3 if farmer did not use with all conditions present; 4 if incorporated organic into flooded material | 1.8686<br>(1.4116) |
| 9. Land Use | Total paddy land size wet season (in hectares) | |
| landuse_v1 | Land use for non-fragrant white rice | 0.515<br>(1.8774) |
| landuse_v2 | Land use for SKO | 0.9257<br>(2.3462) |
| landuse_v3 | Land use for Jasmine rice, Phka Rumduol and other types | 1.2566<br>(1.6241) |
| 10. Buyers of Rice | To whom do you normally sell your rice? (Select all that apply) | |
| offtaker_amru | 1 if Amru Rice Company; 0 otherwise | 0.178<br>(0.3828) |
| offtaker_coop | 1 if cooperative; 0 otherwise | 0.072<br>(0.2587) |
| offtaker_farmers | 1 if other farmers; 0 otherwise | 0.0537<br>(0.2255) |
| offtaker_localmarket | 1 if local market; 0 otherwise | 0.0226<br>(0.1487) |
| offtaker_middlemen | 1 if middlemen; 0 otherwise | 0.7669<br>(0.4231) |
| offtaker_millers | 1 if other millers; 0 otherwise | 0.0523<br>(0.2227) |
| offtaker_other | Other | 0.0607<br>(0.239) |
| 11. Gender: Ownership/Decision making/ Participation | Who does the following in your household? (Select one per row) | |
| ownership_bankaccount | Name on bank account; 1 if men only; 2 if mostly men; 3 if men and women equally; 4 if mostly women; 5 if women only; 6 if not applicable | 3.3051<br>(1.1702) |
| ownership_hhexpenditure | Manage household expenditure; 1 if men only; 2 if mostly men; 3 if men and women equally; 4 if mostly women; 5 if women only; 6 if not applicable | 3.8277<br>(0.9372) |
| ownership_land | Owner of farm land; 1 if men only; 2 if mostly men; 3 if men and women equally; 4 if mostly women; 5 if women only; 6 if not applicable | 2.9449<br>(0.7397) |
| ownership_loan | Name on loan agreement; 1 if men only; 2 if mostly men; 3 if men and women equally; 4 if mostly women; 5 if women only; 6 if not applicable | 3.3065<br>(1.0851) |
| ownership_meetings | Participate in farm related meetings; 1 if men only; 2 if mostly men; 3 if men and women equally; 4 if mostly women; 5 if women only; 6 if not applicable | 3.5791<br>(1.3553) |
| 12. Supply Source of Pesticides | Where did you get your pesticides from? (Select all that apply) | |

| | | |
|---|---|---|
| pest_provider_Cooperativ | 1 if cooperative; 0 otherwise | 0.178<br>(0.3828) |
| pest_provider_Directly_f | 1 if directly from rice company or processor;<br>0 otherwise | 0.0028<br>(0.0531) |
| pest_provider_I_did_not | 1 if farmer did not use pesticides;<br>0 otherwise | 0.0749<br>(0.2633) |
| pest_provider_Local_shop | 1 if local shop; 0 otherwise | 0.8588<br>(0.3485) |
| pest_provider_Others__SP | Others | 0.0085<br>(0.0917) |
| 13. Prices | What price per kilo, tons or sacks did<br>you receive in the 2021 wet season harvest?<br>(In Cambodian Riel) | |
| price_v1_kg | Price per kilo of non-fragrant white rice | 22.0339<br>(133.0805) |
| price_v1_sack | Price per sack of non-fragrant white rice | 1488.5593<br>(10435.7709) |
| price_v1_ton | Price per ton of non-fragrant white rice | 84576.2712<br>(256264.0216) |
| price_v2_kg | Price per kilo of SKO | 29.3362<br>(168.4163) |
| price_v2_sack | Price per sack of SKO | 84.7458<br>(2254.9381) |
| price_v2_ton | Price per ton of SKO | 165658.1921<br>(363160.746) |
| price_v3_kg | Price per kilo of Jasmine, Phka Rumduol,<br>and other types | 77.5<br>(280.0078) |
| price_v3_sack | Price per sack of Jasmine, Phka Rumduol,<br>and other types | 5566.3842<br>(19848.0038) |
| price_v3_ton | Price per ton of of Jasmine, Phka Rumduol,<br>and other types | 455992.9379<br>(526616.8085) |
| 14. Rice Sales | How many kilos, ton or sacks of paddy rice<br>did you produce in the 2021 wet season (wet volume)? | |
| production_v1_kg | Kilograms of non-fragrant white rice | 303.6441<br>(2606.168) |
| production_v1_kgsack | Kilograms/sack of non-fragrant white rice | 2.1681<br>(12.7175) |
| production_v1_sack | Number of sacks of non-fragrant white rice | 0.8475<br>(5.8093) |
| production_v1_ton | Tons of non-fragrant white rice | 1.1979<br>(6.2193) |
| production_v2_kg | Kilograms of Sen Kra Oub | 747.2006<br>(5559.4661) |
| production_v2_kgsack | Kilograms/sack of Sen Kra Oub | 0.3249<br>(5.105) |
| production_v2_sack | Number of sacks of Sen Kra Oub | 0.0621<br>(1.1972) |
| production_v2_ton | Tons of non-fragrant Sen Kra Oub | 2.661<br>(7.956) |
| production_v3_kg | Kilograms of Jasmine rice, Phka Rumduol<br>and other types | 314.7638<br>(1541.158) |
| production_v3_kgsack | Kilograms/sack of Jasmine rice, Phka Rumduol<br>and other types | 8.0042<br>(22.018) |
| production_v3_sack | Number of sacks of Jasmine rice, Phka Rumduol<br>and other types | 3.1441<br>(11.8459) |

| | | |
|---|---|---|
| production_v3_ton | Tons of Jasmine rice, Phka Rumduol and other types | 2.0243 (3.6389) |
| 15. Supply Source of Seeds | Where did you get most of your seeds for 2021 rice production? | |
| seed | 1 if self-saved seeds met criteria for self-daved seeds with quality for a max of crop values; 0 if uncertified seeds or seeds without quality control; 3 if certified seed or seed with quality control suitable for local conditions | 2.1441 (1.1278) |
| seed_provider_AC_or_PDAF | 1 if AC or PDAFF; 0 otherwise | 0.2585 (0.4381) |
| seed_provider_Exchange_w | 1 if exchange with other farmers; 0 otherwise | 0.0508 (0.2198) |
| seed_provider_Local_shop | 1 if local shop; 0 otherwise | 0.0777 (0.2679) |
| seed_provider_Other__SPE | Other | 0.0056 (0.075) |
| seed_provider_Own_saved | 1 if own-saved seeds; 0 otherwise | 0.6059 (0.489) |
| seed_provider_Rice_compa | 1 if rice company or processor; 0 otherwise | 0.0014 (0.0376) |
| Seeding Practices | Which of the following seed planting practices did you use? (single choice) | |
| seedplant_Direct_see | 1 if direct seeding or hand spraying seeds; 0 otherwise | 0.9379 (0.2416) |
| seedplant_Drum_seedi | 1 if drum seeding; 0 otherwise | 0.0155 (0.1238) |
| seedplant_Planting_m | 1 if planting machine; 0 otherwise | 0.0071 (0.0838) |
| seedplant_Transplant | 1 if transplanting; ; 0 otherwise | 0.0395 (0.195) |
| 16. Rice Sales Wet Season 2021 | How many kilos, tons, or sacks of paddy rice did you sell or trade in the 2021 wet season harvest? | |
| sold_v1_kg | Kilograms of non-fragrant white rice | 274.8729 (2510.5906) |
| sold_v1_kgsack | Kilograms/sack of non-fragrant white rice | 1.6667 (11.4866) |
| sold_v1_sack | Number of sacks of non-fragrant white rice | 0.5385 (4.5124) |
| sold_v1_ton | Tons of non-fragrant white rice | 0.9866 (5.3616) |
| sold_v2_kg | Kilograms of Sen Kra Oub | 697.0353 (5260.7756) |
| sold_v2_kgsack | Kilograms/sack of Sen Kra Oub | 0.226 (4.3801) |
| sold_v2_sack | Number of sacks of Sen Kra Oub | 0.0424 (1.1275) |
| sold_v2_ton | Tons of non-fragrant Sen Kra Oub | 2.4018 (7.4136) |
| sold_v3_kg | Kilograms of Jasmine rice, Phka Rumduol and other types | 217.7455 (1297.669) |
| sold_v3_kgsack | Kilograms/sack of Jasmine rice, Phka Rumduol and other types | 213.2528 (3903.4633) |
| sold_v3_sack | Number of sacks of Jasmine rice, Phka Rumduol and other types | 1.5042 (7.0961) |

| | | |
|---|---|---|
| sold_v3_ton | Tons of Jasmine rice, Phka Rumduol and other types | 1.4068<br>(2.9733) |
| sold_everything | Are you able to sell all the rice you want to sell?<br>1. Yes, all the time; 2. Most of the time;<br>3. Sometimes; 4. Seldom | 2.1271<br>(1.1211) |
| 16. AC Support Type | What kinds of support do you receive from<br>the cooperative/ farmer group? | |
| supp_type_Other__SPECIFY_ | Other | 0.0014<br>(0.0376) |
| supp_type_Supplying_good_ri | 1 if supplying good rice seeds; 0 otherwise | 0.0579<br>(0.2337) |
| supp_type_Training_of_clima | 1 if training of climate smart agriculture practices;<br>0 otherwise | 0.1088<br>(0.3116) |
| supp_type_Training_on_AC__A | 1 if training on Agriculture Cooperative management;<br>0 otherwise | 0.0212<br>(0.1441) |
| supp_type_Training_on_farm | 1 if training on farm management;<br>0 otherwise | 0.0339<br>(0.1811) |
| supp_type_Training_on_farmi | 1 if training on farming techniques in general;<br>0 otherwise | 0.0508<br>(0.2198) |
| supp_type_Training_on_finan | 1 if training on financial literacy;<br>0 otherwise | 0.0636<br>(0.2441) |
| supp_type_Training_on_SRP_s | 1 if training on SRP standards in particular;<br>0 otherwise | 0.0678<br>(0.2516) |
| 17. Training Provider | Who provided the training? (Select all that apply) | |
| train_provider_bank | 1 if bank; 0 otherwise | 0.0508<br>(0.2198) |
| train_provider_CIRD | 1 if CIRD; 0 otherwise | 0.3093<br>(0.4625) |
| train_provider_Coop | 1 if cooperative; 0 otherwise | 0.5579<br>(0.497) |
| train_provider_dremember | 1 if don't remember; 0 otherwise | 0.0339<br>(0.1811) |
| train_provider_Gov | 1 if government; 0 otherwise | 0.2048<br>(0.4038) |
| train_provider_NGO | 1 if NGO; 0 otherwise | 0.3658<br>(0.482) |
| train_provider_Other | Other | 0.0311<br>(0.1736) |
| train_provider_psa | 1 if private sector agent; 0 otherwise | 0.2486<br>(0.4325) |
| 18. Participation in training | Have your or any in your household<br>attended any of the following training<br>in the last 2 years?<br>(Select single option for each topic) | |
| training_attend_ac | AC management; 1. Yes, general practices;<br>2. Yes, SRP standards; 3. No; 4. Do not know | 1.774<br>(0.81) |
| training_attend_fertilizing | Fertilizing; 1. Yes, general practices;<br>2. Yes, SRP standards; 3. No; 4. Do not know | 2.1031<br>(1.054) |
| training_attend_finmanage | Family financial management;<br>1. Yes, general practices; 2. Yes, SRP standards;<br>3. No; 4. Do not know | 1.6624<br>(0.8174) |
| training_attend_goals | Set and implement family goals;<br>1. Yes, general practices; 2. Yes, SRP standards;<br>3. No; 4. Do not know | 1.6921<br>(0.8196) |
| training_attend_harv | Harvesting; 1. Yes, general practices;<br>2. Yes, SRP standards; 3. No; 4. Do not know | 2.2599<br>(1.0718) |

| | | |
|---|---|---|
| training_attend_irrigation | Water management and irrigation; 1. Yes, general practices; 2. Yes, SRP standards; 3. No; 4. Do not know | 3.1864 (1.9619) |
| training_attend_landprep | Land preparation and planning; 1. Yes, general practices; 2. Yes, SRP standards; 3. No; 4. Do not know | 3.0664 (1.9228) |
| training_attend_other | Other; 1. Yes, general practices; 2. Yes, SRP standards; 3. No; 4. Do not know | 2.3008 (0.5764) |
| training_attend_pest | Pest and disease management; 1. Yes, general practices; 2. Yes, SRP standards; 3. No; 4. Do not know | 2.1864 (1.0831) |
| training_attend_postharv | Post harvesting; 1. Yes, general practices; 2. Yes, SRP standards; 3. No; 4. Do not know | 2.3249 (1.0646) |
| training_attend_rotation | Crop cover/rotation; 1. Yes, general practices; 2. Yes, SRP standards; 3. No; 4. Do not know | 2.6045 (1.0804) |
| training_attend_seedproduction | Good seed production; 1. Yes, general practices; 2. Yes, SRP standards; 3. No; 4. Do not know | 1.6186 (0.7952) |
| training_attend_seedvar | Seed variety; 1. Yes, general practices; 2. Yes, SRP standards; 3. No; 4. Do not know | 2.1525 (1.0527) |
| training_attend_traceability | Traceablility system for paddy quality; 1. Yes, general practices; 2. Yes, SRP standards; 3. No; 4. Do not know | 1.7797 (0.785) |
| 19. Training Provider for Specific Tranings | Who provided the training? (Select all that apply) | |
| training_provider_ac | Agricultural Cooperative management; 1 if CIRD; 2 if NGO; 3 if bank; 4 if cooperative; 5 if private sector agent; 6 if government; 7 if other; 8 if don't remember; 0 if no training | 0.9576 (1.3853) |
| training_provider_fertilizing | Fertilizing; 1 if CIRD; 2 if NGO; 3 if bank; 4 if cooperative; 5 if private sector agent; 6 if government; 7 if other; 8 if don't remember; 0 if no training | 1.6398 (1.7132) |
| training_provider_finmanage | Financial management; 1 if CIRD; 2 if NGO; 3 if bank; 4 if cooperative; 5 if private sector agent; 6 if government; 7 if other; 8 if don't remember; 0 if no training | 1.1017 (1.3964) |
| training_provider_goals | Set and implement family goals; 1 if CIRD; 2 if NGO; 3 if bank; 4 if cooperative; 5 if private sector agent; 6 if government; 7 if other; 8 if don't remember; 0 if no training | 1.0551 (1.3884) |
| training_provider_harv | Harvesting; 1 if CIRD; 2 if NGO; 3 if bank; 4 if cooperative; 5 if private sector agent; 6 if government; 7 if other; 8 if don't remember; 0 if no training | 1.3573 (1.5491) |
| training_provider_irrigation | Irrigation; 1 if CIRD; 2 if NGO; 3 if bank; 4 if cooperative; 5 if private sector agent; 6 if government; 7 if other; 8 if don't remember; 0 if no training | 1.3573 (1.5582) |
| training_provider_landprep | Land preparation and planning; 1 if CIRD; 2 if NGO; 3 if bank; 4 if cooperative; 5 if private sector agent; 6 if government; 7 if other; 8 if don't remember; 0 if no training | 1.4124 (1.5256) |
| training_provider_oth | Other; 1 if CIRD; 2 if NGO; 3 if bank; 4 if cooperative; 5 if private sector agent; 6 if government; 7 if other; 8 if don't remember; 0 if no training | 0.2345 (1.0803) |
| training_provider_pest | Pest and disease management; 1 if CIRD; 2 if NGO; 3 if bank; 4 if cooperative; 5 if private sector agent; 6 if government; 7 if other; 8 if don't remember; 0 if no training | 1.5085 (1.6526) |

| | | |
|---|---|---|
| training_provider_postharv | Post harvesting; 1 if CIRD; 2 if NGO; 3 if bank; 4 if cooperative; 5 if private sector agent; 6 if government; 7 if other; 8 if don't remember; 0 if no training | 1.2811 (1.5142) |
| training_provider_rotation | Crop cover/rotation; 1 if CIRD; 2 if NGO; 3 if bank; 4 if cooperative; 5 if private sector agent; 6 if government; 7 if other; 8 if don't remember; 0 if no training | 0.7811 (1.3742) |
| training_provider_seedproduction | Good seed production; 1 if CIRD; 2 if NGO; 3 if bank; 4 if cooperative; 5 if private sector agent; 6 if government; 7 if other; 8 if don't remember; 0 if no training | 1.178 (1.4476) |
| training_provider_seedvar | Seed variety; 1 if CIRD; 2 if NGO; 3 if bank; 4 if cooperative; 5 if private sector agent; 6 if government; 7 if other; 8 if don't remember; 0 if no training | 1.5042 (1.6578) |
| training_provider_traceability | Traceability system for paddy quality; 1 if CIRD; 2 if NGO; 3 if bank; 4 if cooperative; 5 if private sector agent; 6 if government; 7 if other; 8 if don't remember; 0 if no training | 0.9407 (1.3908) |
| 20. Reasons for no training participation | Why did you not attend this training? | |
| training_whyno_ac | AC management; 1 = I did not have time; 2 = I was not interested; 3 = I was not aware about the training; 4 = Such training was not available; 5 = Other (SPECIFY) | 1.0311 (1.6984) |
| training_whyno_fertilizing | Fertilizing; ; 1 = I did not have time; 2 = I was not interested; 3 = I was not aware about the training; 4 = Such training was not available; 5 = Other (SPECIFY) | 0.4506 (1.239) |
| training_whyno_finmanage | Financial management; 1 = I did not have time; 2 = I was not interested; 3 = I was not aware about the training; 4 = Such training was not available; 5 = Other (SPECIFY) | 0.7655 (1.5328) |
| training_whyno_goals | Set and implement family goals; 1 = I did not have time; 2 = I was not interested; 3 = I was not aware about the training; 4 = Such training was not available; 5 = Other (SPECIFY) | 0.8065 (1.5609) |
| training_whyno_harv | Harvesting; 1 = I did not have time; 2 = I was not interested; 3 = I was not aware about the training; 4 = Such training was not available; 5 = Other (SPECIFY) | 0.6158 (1.4239) |
| training_whyno_irrigation | Irrigation; 1 = I did not have time; 2 = I was not interested; 3 = I was not aware about the training; 4 = Such training was not available; 5 = Other (SPECIFY) | 0.6328 (1.4383) |
| training_whyno_landprep | Land preparation and planting; 1 = I did not have time; 2 = I was not interested; 3 = I was not aware about the training; 4 = Such training was not available; 5 = Other (SPECIFY) | 0.589 (1.3938) |
| training_whyno_oth | Other; 1 = I did not have time; 2 = I was not interested; 3 = I was not aware about the training; 4 = Such training was not available; 5 = Other (SPECIFY) | 1.8475 (1.8545) |

| | | |
|---|---|---|
| training_whyno_pest | Pest and disease management; 1 = I did not have time; 2 = I was not interested; 3 = I was not aware about the training; 4 = Such training was not available; 5 = Other (SPECIFY) | 0.5028 (1.2968) |
| training_whyno_postharv | Post harvesting; 1 = I did not have time; 2 = I was not interested; 3 = I was not aware about the training; 4 = Such training was not available; 5 = Other (SPECIFY) | 0.7203 (1.5159) |
| training_whyno_rotation | Crop cover/rotation; 1 = I did not have time; 2 = I was not interested; 3 = I was not aware about the training; 4 = Such training was not available; 5 = Other (SPECIFY) | 1.3842 (1.8175) |
| training_whyno_seedproduction | Good seed production; 1 = I did not have time; 2 = I was not interested; 3 = I was not aware about the training; 4 = Such training was not available; 5 = Other (SPECIFY) | 0.7684 (1.5314) |
| training_whyno_seedvar | Seed variety; 1 = I did not have time; 2 = I was not interested; 3 = I was not aware about the training; 4 = Such training was not available; 5 = Other (SPECIFY) | 0.5706 (1.3764) |
| training_whyno_traceability | Traceability system for paddy quality; 1 = I did not have time; 2 = I was not interested; 3 = I was not aware about the training; 4 = Such training was not available; 5 = Other (SPECIFY) | 1.2147 (1.7987) |
| 21. Variety | What rice varieties did you grow on this farm in wet season 2021? (Select all that apply) | |
| variety_othertype | 1 if fragrant rice other type; 0 otherwise | 0.1427 (0.35) |
| variety_rumdeng | 1 if Phka Rumdeng; 0 otherwise | 0.0056 (0.075) |
| variety_rumduol | 1 if Phka Rumduol; 0 otherwise | 0.5847 (0.4931) |
| variety_sko | 1 if Sen Kra Oub; 0 otherwise | 0.2203 (0.4148) |
| variety_sticky | 1 if sticky rice; 0 otherwise | 0.0028 (0.0531) |
| variety_white | 1 if non-fragrant white rice; 0 otherwise | 0.1667 (0.3729) |

Figure 1: Visualization Map of the Predictive Results for NN Model for Cambodia

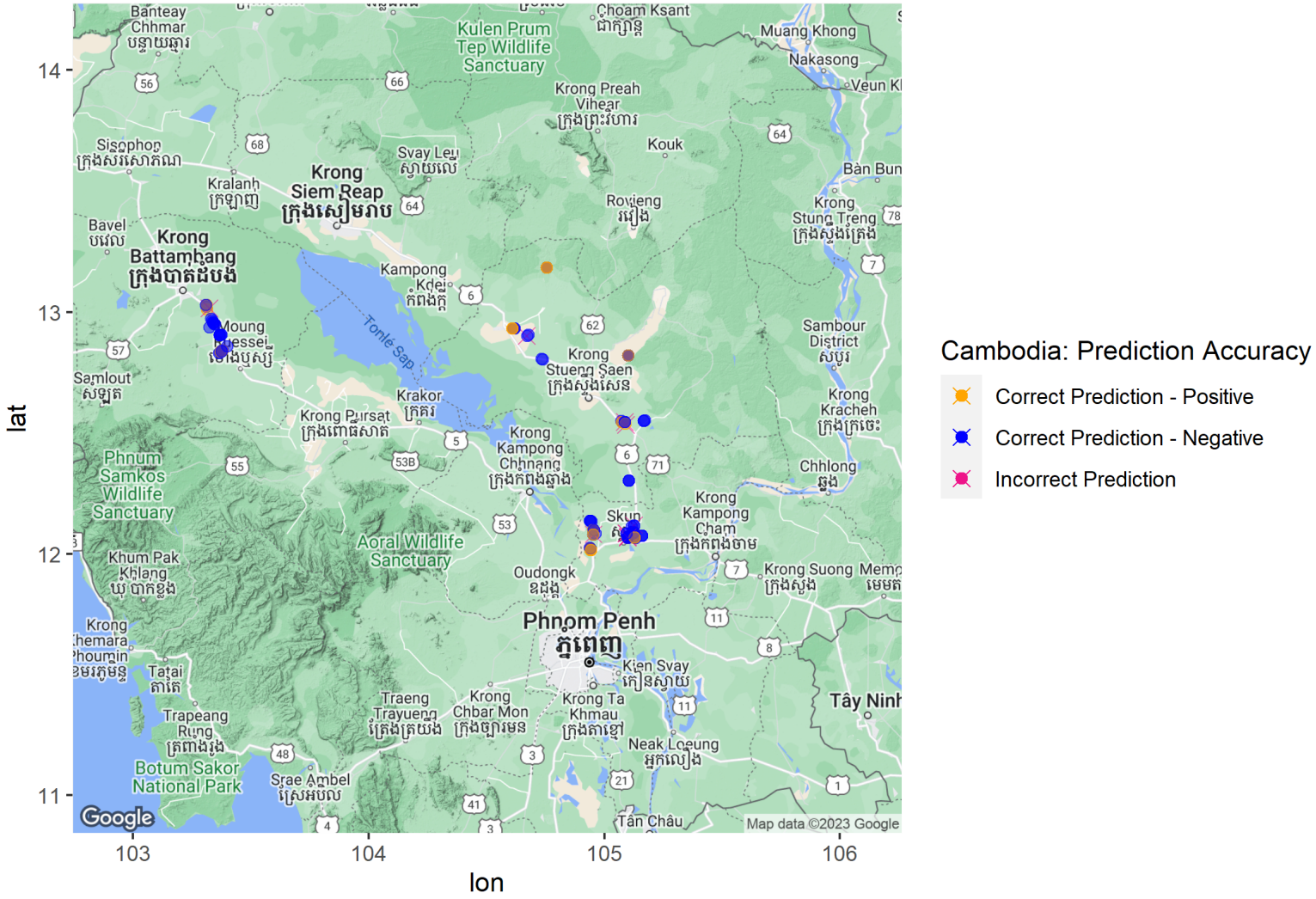


Figure 2: Visualization Map of the Predictive Results for NN Model for Ethiopia

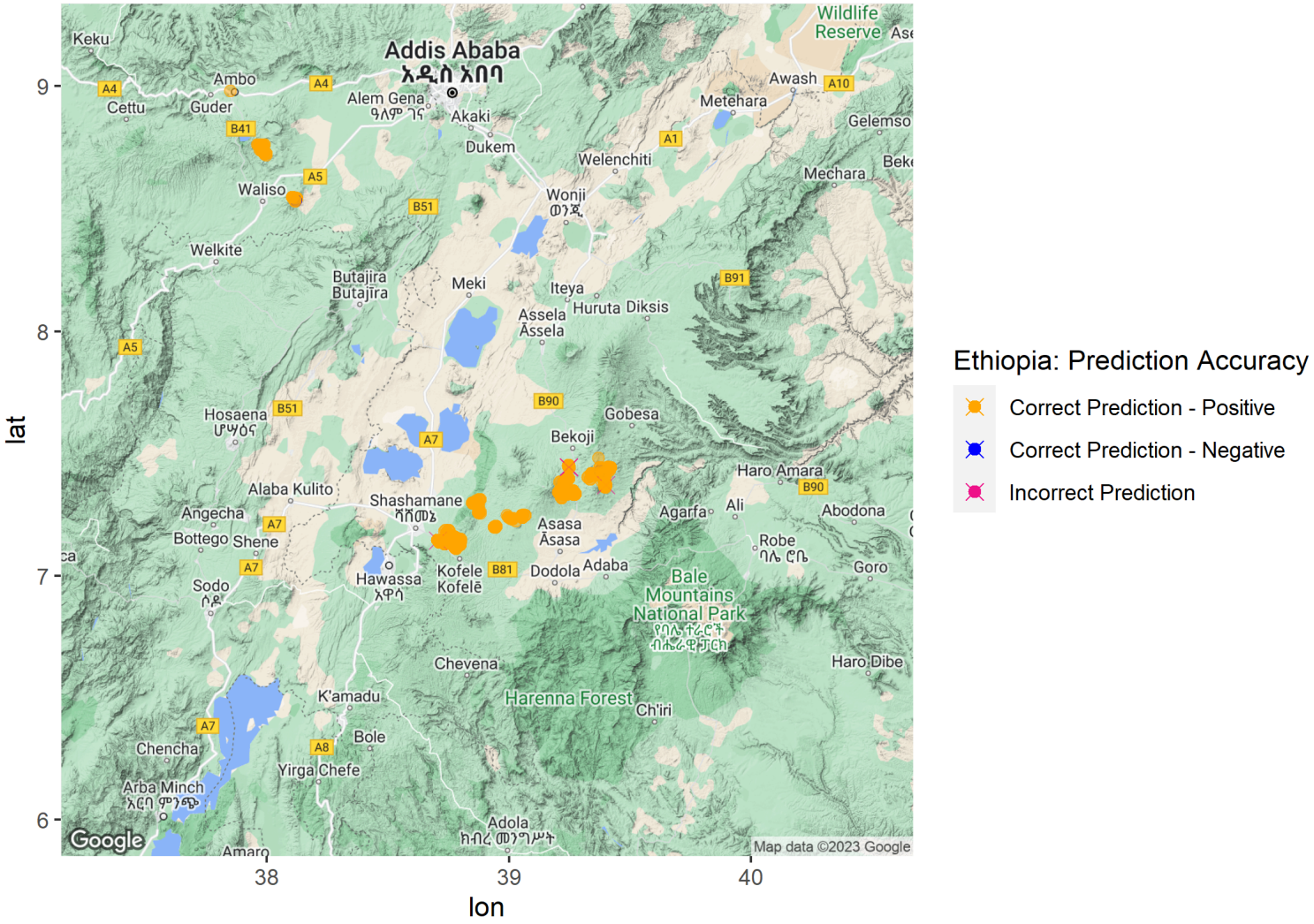


Figure 3: RF Feature Importance for Top 30 Covariates with Cambodian Amru (Rice) Dataset

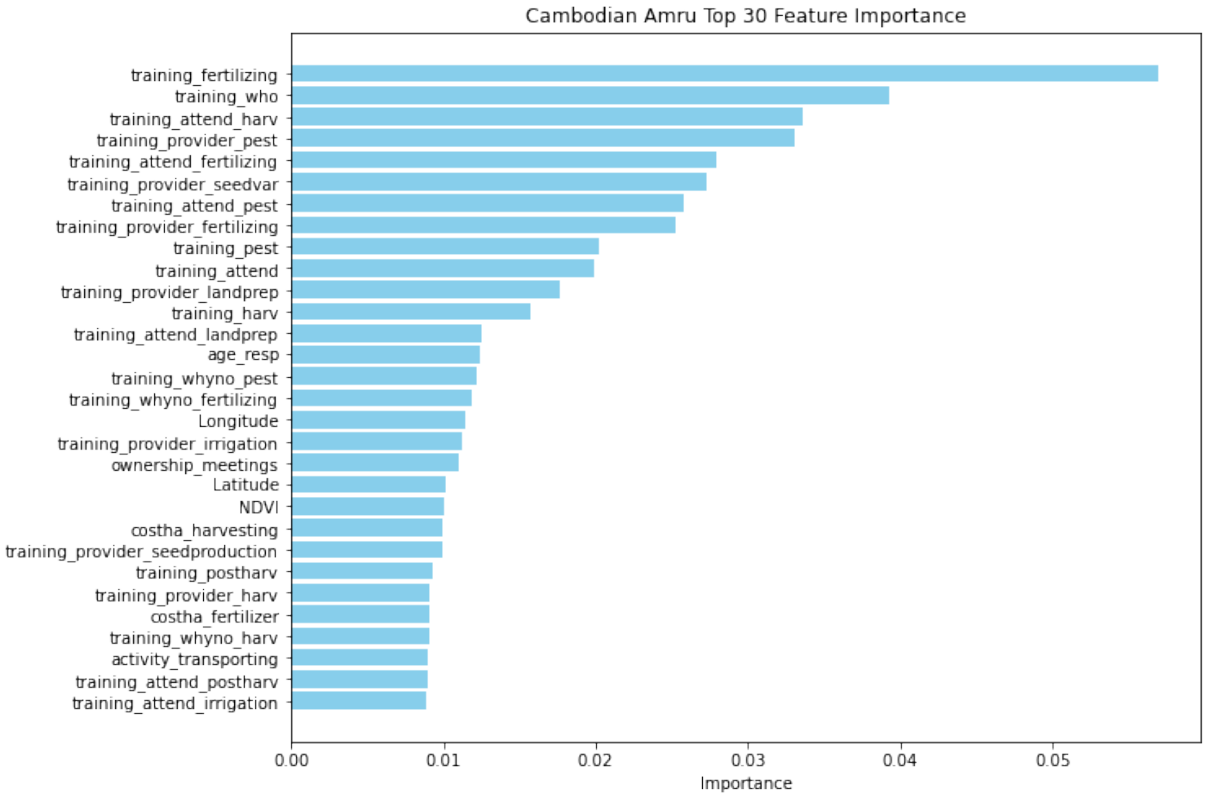



Figure 4: RF Feature Importance for Top 30 Covariates with Ethiopian Heineken (Barley) Dataset

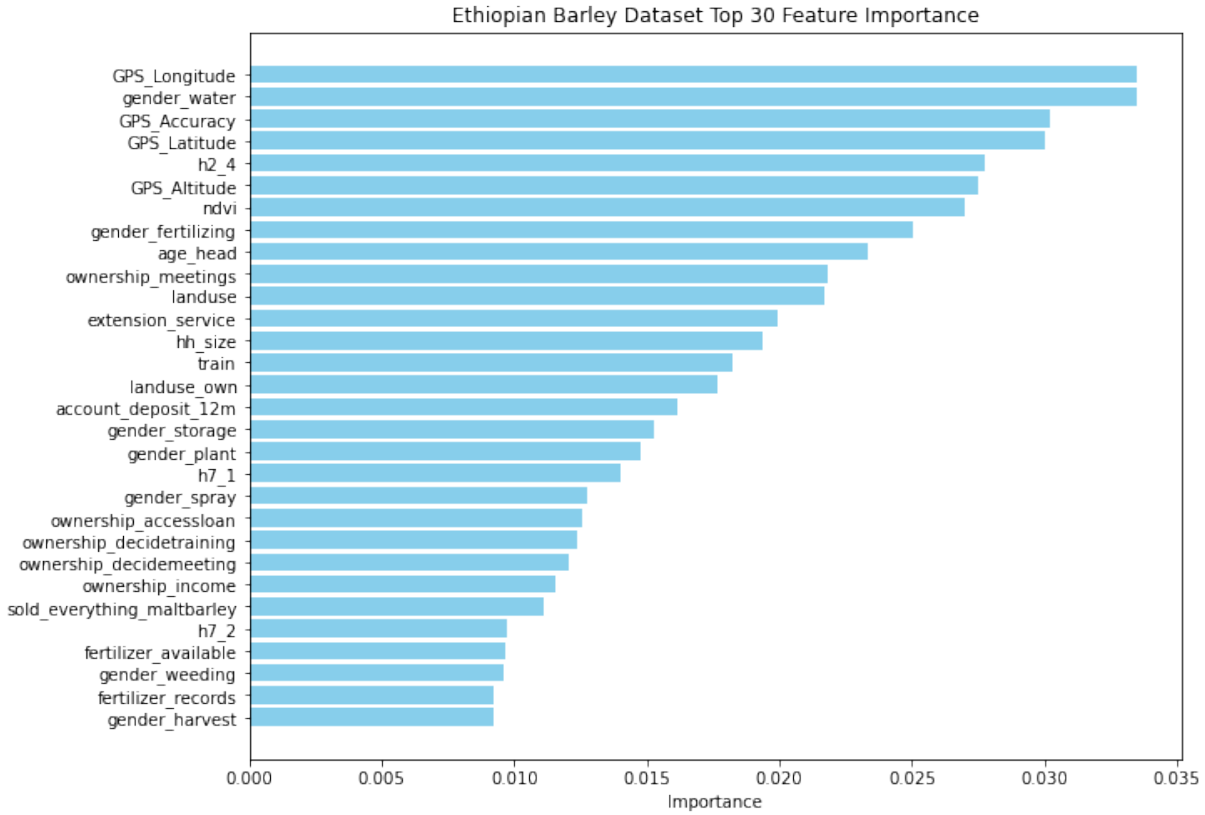


Figure 5: RF Feature Importance for Top 30 Covariates with Cross-Country Dataset

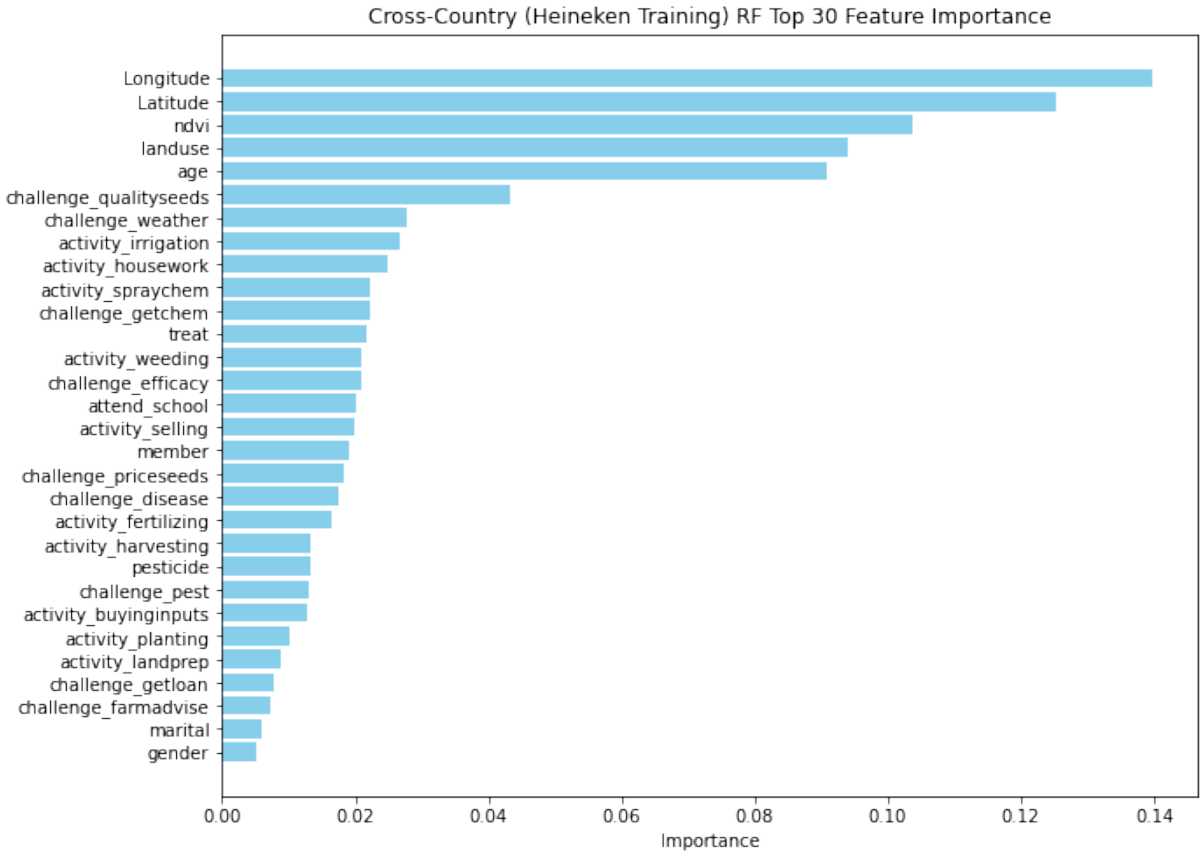
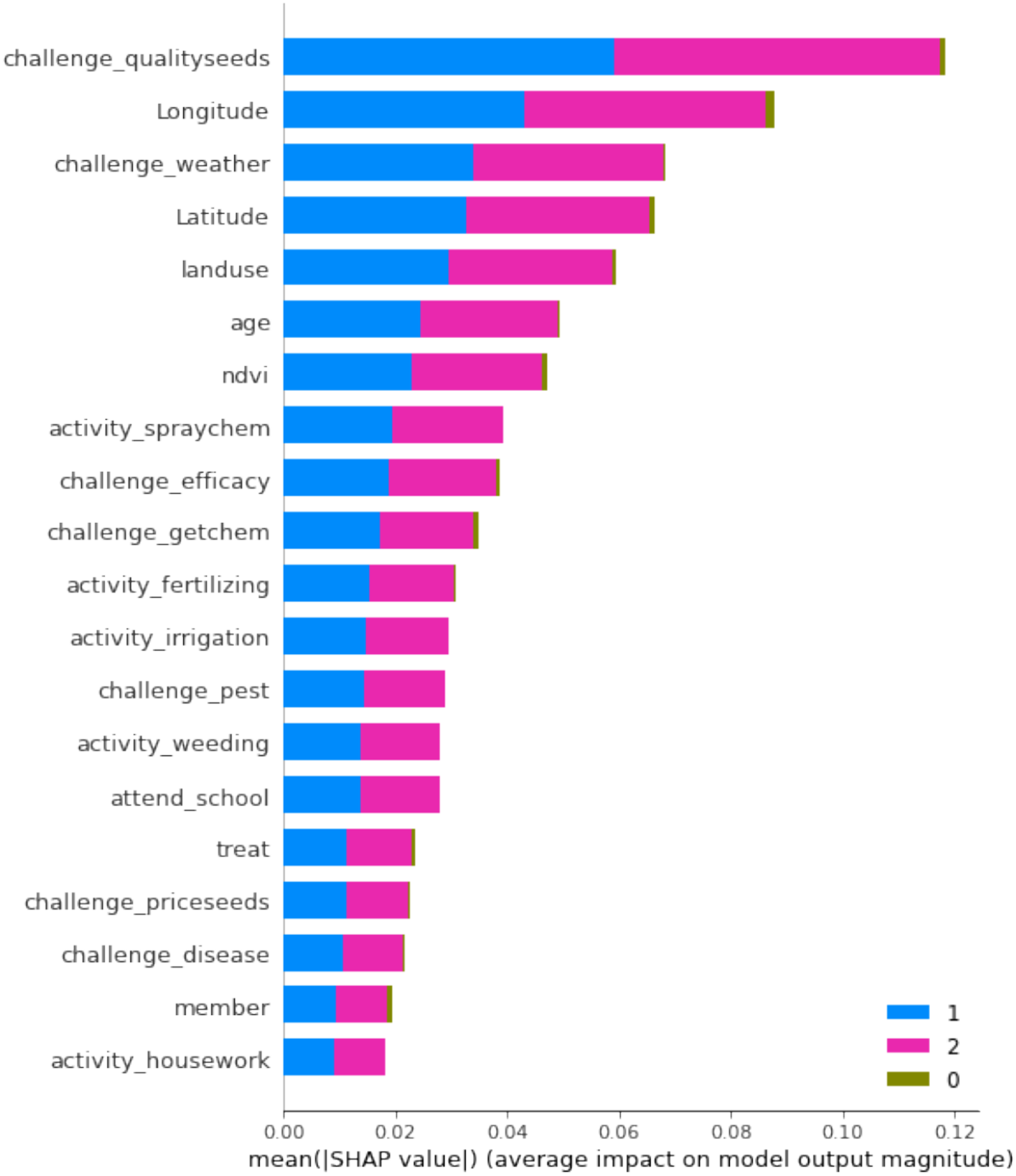


Figure 6: Shapley Values with Cross-Country Dataset




A value of 1 (blue) represents the class of farmers who do not need additional training on GAP (a proxy for long term adoption) and a value of 2 (red) represents the class of farmers who do need additional training on GAP.