Master's programme: Applied Data Science

Utrecht University, the Netherlands

MSc Thesis Marion Späth (2772981)

TITLE: Machine Learning for Welfare Prediction: Insights from IBLI Insurance Data in Southern Ethiopia

July 2025

Supervisor 1:

Dr. Mahdi Shafiee Kamalabad

Supervisor 2:

Samin Nikkhah Bahrami

Word count: 9222

Abstract

This study investigates to what extent machine learning (ML) can be used to predict welfare outcomes using household and consumer characteristics in the context of IBLI insurance in Southern Ethiopia. By focusing on welfare outcomes rather than simple insurance uptake rates and thereby allowing for varying decision qualities, this study closes an important gap in the literature. The aim is to identify relevant variables and subgroups for which targeted interventions can be designed by policymakers. Administrative insurance and survey panel data collected over three rounds in a Randomized Control Trial from households of goat and cattle herders in the rural Borena zone in Southern Ethiopia is used, and four ML models (i.e., Elastic Net, Random Forest, CatBoost, and SVR) are run with a focus on prediction accuracy and predictor stability across models. Results show extremely weak prediction accuracy and an explained variances of no more than 1.5% in the best performing model despite using a large set of predictors. Effect sizes are very weak, but in relative terms trust in the insurance, the non-purchase reason, and the main information source are consistently the strongest predictors across models, both for goat and cattle herders. More specifically, external constraints, low trust, and interpersonal contacts as the main information source tend to have welfare reducing effects. Policymakers are advised to explore those three factors in more depth and to identify other variables such as cognitive differences that are more closely related to welfare.

Content

1. Introduction

Index-Based Livestock Insurance (IBLI) aims to protect insured pastoralist households in rural areas of developing countries from the adverse consequences of natural disasters such as floods or droughts (Barrett et al., 2024; Chantarat et al., 2013). In contrast to traditional insurance products, where losses are reimbursed on an individual claim basis, IBLI relies on remote sensing techniques and monitors indices assumed to be indicative of the occurrence of natural disasters (Chantarat et al., 2017). For example, the Normalized Difference Vegetation Index (NDVI) is a common measure of green vegetation cover, and low values relative to historical data records suggest droughts. Indemnity payments to insured households are then automatically triggered if the index strikes a certain contractually determined threshold (Chantarat et al., 2017).

Importantly, IBLI is always accompanied by a basis risk which is the risk that the reimbursement will not be paid despite the occurrence drought due to a faulty or insufficient index value (Chantarat et al., 2017). For example, it could be that a drought occurs and causes losses to insured households but is not picked up by the index (e.g., due to measurement errors) or remains below its threshold. In that case, insured household would not be compensated and would additionally lose the premium they paid for the product. Thus, the effective and accurate protection of households depends on the reliability of the chosen index. Still, IBLI has the advantage of being objectively verifiable and substantially reduces transaction and administration costs making the insurance more easily available to pastoralists with low (financial) literacy and resources (Chantarat et al., 2017; Zewdie et al., 2020). Indeed, empirical research shows that IBLI coverage is linked to reduced conflict (Gehring and Schaudt, 2024), increased well-being which persists even when payments fails (Tafere et al., 2018), improved food safety (Taye, 2023), and faster recovery from natural shocks (Bertram-Huemmer and Kraehnert, 2017).

However, despite these positive outcomes, uptake rates are generally low (Takahashi et al., 2019). Importantly though, insurance uptake does not always equal welfare maximisation for individuals (Chantarat et al., 2017; Harrison et al., 2020). Whether the purchase of an insurance product is welfare maximising for an individual depends on many factors such as their risk-aversion and personal assets and is surrounded by uncertainty (Chantarat et al., 2017; Handel et al., 2020; Harrison et al., 2020). Research shows that a fully informed decision-making and a thorough understanding of the alternatives and consequences are difficult for the average consumer to understand, often leading to suboptimal choices in which individuals forgo welfare and could have instead benefitted from a different decision (Handel et al., 2020; Harrison et al., 2020). Additionally, behavioural interventions do not always lead to improved decision-making in terms of welfare (Harrison et al., 2020). For example, Harrison et al. (2020) show that interventions aiming to improve individuals' understanding through the provision of information and practice increase insurance take-up. Crucially however, they at the same time, on average, decrease welfare due to excessive purchases by individuals who would have been better off without the purchase. Research also suggests that events and interventions can have heterogeneous effects across specific subgroups (e.g., gender or education) (Athey et al., 2025; Athey et al., 2023; Harrison et al., 2020) further complicating the effectiveness of generic interventions aimed at welfare maximisation. Automated tools taking into account consumers' personal characteristics and other context information to offer optimal, tailored advise as well as subgroup specific interventions could be one solution to improve financial decisions and increase overall welfare (Harrison et al., 2020). Hence, rather than designing interventions that target a general uptake of IBLI, thereby decreasing welfare for many, it is important to identify the characteristics of who benefits and who loses from IBLI (non-) purchase to be able to design targeted intervention for identified subgroups.

So far, most work on differential welfare effects is done outside of the IBLI context through experiments with students (Harrison et al., 2020) or using data from industrialised countries (Handel et al., 2020). Thus, to get a first understanding of how consumer and household characteristics are related to welfare from IBLI specifically, research into their relationships with uptake rates can be helpful, nonetheless. Findings are often contradictory (Jensen et al., 2018; Takahashi et al., 2015; Takahashi et al., 2019, Taye, 2023). For example, Taye (2023) descriptively reports that insured households, in comparison to uninsured households, are mostly male, above the age of 40, have multiple sources of income, and a higher socio-economic status. Similarly, Takahashi et al. (2019) find income and education to be positively related to IBLI purchase. Yet, Takahashi et al. (2015) report a negative association between education and IBLI uptake. Jensen et al. (2018), using more sophisticated regression techniques and longitudinal data, report IBLI demand and household characteristics to be largely unrelated, showing no consistent relationship "between demand for IBLI and gender, age of household head, education, income, asset wealth, or ratio of income from livestock, an indicator of the relative risk that drought could pose to the household" (p. 184). The author finds that only households with high levels of participation in social groups show more inclination towards IBLI uptake (Jensen et al., 2018).

Research further shows that, in contrast to theoretical expectations, which typically predict highly risk-averse individuals to purchase insurance more often than less risk-averse individuals, risk aversion in the context of IBLI products is inversely related to product uptake. That is, households with low risk aversion show higher rates of taking up IBLI than households with high risk aversion, which could indicate poor understanding of the product (Takahashi et al., 2019). Yet, while improving product understanding, behavioural interventions do not seem to be causally associated with uptake rates (Harrison et al., 2020, Takahashi et al., 2015). In contrast, random distribution of discount coupons shows significant and sustainable increases in IBLI uptake (Takahashi et al., 2015; Takahashi et al., 2019). This is important, as research suggests "learning-by-doing effects" (Takahashi et al., 2019, p. 24), showing that past IBLI uptake strongly predicts future uptake. Taken together, initial financial incentives could be a key mechanism for increasing the rate of insured households in the long run. Still, further research is needed to identify and understand how household characteristics are related to IBLI uptake, or as elaborated above welfare from IBLI.

This study investigates the household and consumer characteristics that might predict specifically the welfare outcomes from insurance decisions. This can help us identify welfare-enhancing and welfare-reducing characteristics. Previously, research on IBLI demand mainly focused on insurance purchase but did not consider differential welfare effects. An exception is the work by Chantarat et al. (2017) which demonstrates that heterogenous welfare effects are also relevant for the context of IBLI insurance products. The authors demonstrate that initial herd size combined with the premium loadings for the insurance influence welfare outcomes resulting from IBLI. They further show that IBLI fails to protect the poorest from falling into a poverty trap characterised by a too small, collapsing herd, but is most impactful for vulnerable non-poor households which can be protected from entering said poverty trap (Chantarat et al., 2017). Still, more research is needed to investigate welfare effects beyond the influence of herd size and premium loadings.

The current study aims to fill this gap in the scientific literature by investigating to what extent various household characteristics can predict heterogeneous welfare effects of IBLI. Thus, the contributions are threefold: firstly, the study moves beyond simple uptake rates and specifically focuses on the heterogeneous welfare effects of IBLI insurance. Secondly, the study contributes to a better understanding of which household characteristics are related to welfare from IBLI (non-) purchase. Thirdly, the results can inform policymakers to effectively adapt or develop targeted interventions for IBLI insurance. The following research questions guide the remainder of this study: Which household characteristics predict welfare losses or gains? To what extent can state-of-the-art machine learning techniques improve the predictive accuracy of welfare outcome estimation using these characteristics? To what extent are the identified characteristics reliable across machine learning models?

The following section starts by introducing the dataset and its context. Subsequently, all pre-processing steps will be documented, and the model selection will be explained before presenting the results. The study will end with a discussion, conclusion and reflection on its limitations.

2. Data

To answer the research questions, the study draws on data collected by researchers during a previous study (Barrett et al., 2024). The data focus on an IBLI product that is marketed by the Oromia Insurance Company in the Borana Zone of Southern Ethiopia. This context is suitable as Ethiopia's population consists of around 14% pastoralists who inhabit around 60% of the country and who play an important role in the countries' export of livestock and meat (Zewdie et al., 2020). Further, Ethiopia has been hit by severe droughts in the past years, and such events can be expected to occur more frequently given the global climate-crisis (UN Office for the Coordination of Humanitarian Affairs, 2024). Thus, insights gained from this study could inform future policy development and targeting strategies, thereby helping better to protect pastoralists from the consequences of natural shocks.

In cooperation with the local insurance staff, data was collected within a Randomized Control Trial from 2,416 pastoralists across 240 regional zones over four waves from 2022 to 2023 in six-month intervals. This study uses waves two, three, and four. The data stem both from surveys filled in by the pastoralists as well as from administrative data documenting IBLI sales from the insurance company. Further, data originally was collected from household which held camels, cattle, goat, and sheep. Different types of livestock fulfil different purposes (Bertram-Huemmer and Kraehnert, 2017). For example, camels are often used for transportation and are seen as a long-term asset. In contrast, cattle are typically used for milk production and consumption, while goats provide cashmere wool, the trade of which serves as an additional source of income (Bertram-Huemmer and Kraehnert, 2017). Further, cattle are described as lumpy assets whereas goats are more liquifiable and cash-like (Dercon, 1998; Kaumbata et al., 2020). Thus, it is likely that the characteristics influencing IBLI purchase, and welfare vary across types of livestock. For scoping reasons, this study will analyse data from herders of cattle and goat. Due to their purpose, these livestock are likely to affect herders and their families more severely and imminently in case of droughts than for example camels.

To handle missing data on the independent variables, first forward-fill and backward-fill techniques are applied. Those are simple techniques to handle missing data, but given the short intervals between the waves, it is reasonable to assume that values did not change drastically within a household. An additional, important advantage is, that these methods account for the panel nature of the data. Further, applying simple imputation techniques like these allow our datasets to be compatible with a range of different models. Remaining missing values on the independent variables, which could be neither forward- nor backward imputed as there were no information on any of the rounds, as well as missing values on the dependent variables were removed using listwise deletion. This led to a loss of 58 rows (i.e., person-round combinations, 1.3%) for the goat subsample and a loss of 50 rows for the cattle subsample (1.2%). Thus, the final goat dataset comprises 2235 individual-round combinations, and the final cattle dataset comprises 4142 individual-round combinations.

The following gives an overview of the included variables. Appendix A Table A1-A6 documents in detail how the categorical variables were constructed, and Appendix B table B1 and B2 and presents summary statistics separately for the cattle and goat dataset. The variables included in this study are:

- **Welfare-Difference (Outcome)**: This outcome variable was originally calculated by comparing the expected consumer surplus from the household's actual decision (purchase or non-purchase) with that of a counterfactual scenario in which the household purchases the advised number of IBLI policies for their herd. In essence, the variable captures expected welfare impact of deviating from the (non-) purchase advise. The variable had only negative values and was heavily skewed to the left emphasising the welfare-reducing effect of deviating from the ideal, advised scenario. To approximate a normal distribution which combats biases in the model, the absolute values of the variables are log+1 transformed (see Appendix C figure C1-C4 for the distribution before and after transformation). Subsequently, to account for the panel nature of the data, the residuals of an individual-level random-effect model were extracted. Ultimately, these are used as the outcome in all models. Chapter 2.1 will discuss the transformation process in more depth. Lower values indicate closer alignment with the best-possible, advised scenario and therefore higher welfare, higher values indicate larger deviations from the best-possible, advised scenario and therefore lower welfare. Lower values can be seen as more desirable in terms of welfare maximisation

- **Age**: Indicates the respondents age at the beginning of the data collection. If not available, the information from the next earliest wave is used.
- **Number of Minors (<18 years) in Household**: Numeric and time-variant. Indicates the number of minors, defined as individuals younger than 18 years, living in the household.
- **Number of Adults in Household**: Numeric and time-variant. Indicates the number of adults living in the household.
- **Trust in Insurance / Insurance Promoters, Composite Score:** Numeric and time-variant. Records the mean agreement of four items (each on a scale from 0 to 10) related to trust in the insurance and its promoters. Items ask 1) to what extent the advice of the VIP is important for the household's decision, 2) to what extent the participant believes the VIP gives advise that is in their best interest, 3) how likely it is that the entire IBLI insurance premium in submitted to the insurance company, and 4) how likely it is that the households receives the claim payment that they are entitled to.
- **Afan Oromo Language**: Binary and time-invariant. Indicate whether person can write simple sentences in Afan Oromo (reference category: No).
- **English Language**: Binary and time-invariant. Indicate whether person can write simple sentences in English (reference category: No).
- **Amharic Language**: Binary and time-invariant. Indicate whether person can write simple sentences in Amharic (reference category: No).
- **Agricultural Land**: Binary and time-invariant. Indicates whether household has farmed any agricultural land (reference category: No) (reference category: No).
- **Irrigated Land**: Binary and time-invariant. Indicates whether household has irrigated agricultural land and therefore access to private water sources (reference category: No).
- **Highest Education of Adult**: Categorical and time-invariant. Indicates the respondent's highest education. Like with the age variable, the earliest response is used for this variable. The following categories are used: Never attended school, Elementary, Nursery, Secondary and Higher, and Adult Education (reference category)
- **Household Expends More Than Median**: Binary and time-invariant. Indicates whether household spends above-average expenditure on their herd (reference category: No).

- **Education of Child**: Categorical and time-variant. Indicates the oldest child's highest education. The following categories are used: Elementary (reference category), Never attended school, Secondary, Other
- **Main Activity of Child**: Categorical and time-variant. Indicates the oldest child's main activity. The following categories are used: Student, Working with Livestock, Not working, House/domestic work (reference category), Working
- **Household Description**: Categorical and time-variant. Indicates whether household is either fully settled (reference category), partially settled, or not settled.
- **Main Information Source**: Categorical and time-variant. Indicates what the main source of information was that a household used to get information about the IBLI insurance. The following categories were used: Interpersonal Sources (reference category), Professional and Organizational Sources, Media, Community-Based, and Survey Sources
- **Religion**: Categorical and time-variant. Indicates the respondent's religion. The following categories were used: Traditional/Wakefata, Muslim, Christian (reference category)
- **Owns Phone**: Binary and time-variant. Indicates whether at least one person in a household owns a phone (reference category: No). Can be interpreted as a proxy for access to technology.
- **Household Moved Last 6 Months**: Binary and time-variant. Indicates whether household has moved during the last 6 months (reference category: No).
- **Reason for Non-Purchase**: Categorical and time-variant. Indicates the main reason for not purchasing the IBLI product. The following categories were used: Lack of Awareness or Understanding, Financial, Practical, and Situational Constraints, Bought Insurance (reference category), Distrust or Negative Perception
- **Know VIP**: Binary and time-variant. Indicates whether respondent knew the village insurance promotor (VIP) (reference category: No).
- **Trust in VIP**: Binary and time-variant. Indicates whether respondent trusted VIP (reference category: No).

For all models, numeric variables are standardised to have a mean of zero and a standard deviation of one to make coefficients comparable. Categorical variables are encoded depending on the model's capabilities which will be described in the following section.

The data was split into train and test dataset, stratified on the outcome. 80% of the data (i.e., 3314 cattle / 3523 goat rows) are used for model training, and 20% (i.e., 828 cattle / 880 goat rows) was used for testing. The models will be trained using 10-fold cross-validation to select the best hyperparameters. Cross-validation is preferred over a separate validation set, to avoid idiosyncrasies in the set construction and to exploit as much data as possible for training purposes. 10 folds are a common choice to balance sample size and robust estimations especially for smaller datasets.

2.1. Considerations of the Panel Structure and Preprocessing of the Outcome Variable

As described above, the data used in this study covers three rounds in which households were surveyed in six-month intervals and made a decision to purchase or not purchase IBLI. Thus, the observations across rounds are not independent but are rather organised in a nested structure, meaning the survey rounds are nested within households. In other words, the same households have been repeatedly interviewed. If this is not taken into account when specifying models, their performance and statistical properties (e.g., standard errors and p-values) will be biased and unreliable. Machine learning models, specifically, aim to learn patterns within the data. If observations are not independent, patterns learned by the model can be misleading, representing within-household idiosyncrasies rather than true associations between variables. For example, it could be that there is a household which participated in all three rounds and has high values on two time-invariant variables. This household is represented with three rows in the dataset (i.e., one for each round). If the model does not account for the fact that all three rows in the dataset are from the same household, it will treat it as three independent data point and might mistakenly conclude that there is a pattern (e.g., strong positive correlation) between the two time-invariant variables that have high values since this seems, incorrectly, like a repeating pattern. Thus, the importance of these variables may be falsely inflated. This is especially problematic when trying to generalise the model to new, unseen data as it overfits to the specific household structure of the training data. Methods such as k-fold cross-validation can also fail due to data leakage, as the held-out fold may contain observations that are correlated with those in the training set. It is for these reasons that the data is further pre-processed before being modelled. Specifically, an individual-only random-effect model is estimated. Such a model estimates the part of the variance in the outcome that can be explained by the nested household structure. Results show that 4.78% of cattle data and 6.68% of the goat data can be explained by this structure. While generally low and indicative of most variance remaining between individuals, proper handling will still improve the reliability of later models. The residuals of the random-effect model can be seen as the outcome ‘cleared’ of the nested structure and are therefore extracted and used as the outcome variable in subsequent machine learning models. Appendix E figure E1-E4 reports the relationship between original and residual variable as well as the model coefficients.

3. Methods

3.1. Dimensionality Considerations

The dataset used in this study has a relatively large number of predictors relative to the number of rows. This ratio increases even more when categorical variables with multiple categories are encoded as separate dummies. This makes the distribution in the data space sparser and can introduce the curse of dimensionality, which describes the phenomenon where model accuracy can deteriorate rather than improve as the number of predictors increases, due to noise introduced by these variables (James et al., 2013). There are multiple ways to combat this. First, some categories can be collapsed, thereby reducing the number of variables as less dummies are needed. This has already been taken into consideration when constructing the variables explained above but could not fully solve the issue (see Appendix A table A1-A6). A second, common approach is conducting a principal component analysis (PCA) which aims to reproject the variables on a certain number of orthogonal, uncorrelated components (i.e., axis) with the aim of capturing as much variance as possible.

The number of components is chosen by the researcher, and the decision is driven by both data and domain knowledge. For each component, an eigenvalue can be computed which indicates the amount of variance it captures. The Kaiser criterion was applied to determine the number of components to retain. Components with eigenvalues less than 1 were excluded, as they account for less total variance than an individual observed variable and are therefore considered less informative. In addition to the eigenvalues of the components, their cumulative explained variance should be taken into consideration as few components should ideally capture most of the data's variance. Lastly, factor loadings indicate how strongly variables are related to the various components and are the basis of meaning attribution by the researcher. A clear pattern of factor loadings across components is especially important if the analysis goal requires interpretability of the results.

One of PCA's requirements is that the variables are numeric. However, only three (i.e., age, number of adults, and number of minors) of the 20 predictors used in this study are numeric. As the concept of variance does not apply to categorical variables, PCA cannot be applied here. An alternative is factor analysis for mixed data (FAMD). FAMD allows both numeric and categorical variables, and its results and interpretation are similar to PCA. That is, eigenvalues, cumulative explained variance, and factor loadings are computed for a specified number of components. For this, numeric variables are treated the same way as in PCA. Categorical variables are first encoded into dummy-variables. The resulting columns are then centered to have a mean of zero and subsequently multiplied by the square-root of the category's frequency. This process is needed to ensure that both frequent and rare categories as well as the numeric variables are equally considered for the computation of the components. After this transformation, standard PCA is performed on the data.

The cattle subsample has been used to explore the potential benefit of FAMD for dimensionality reduction in this study. Results (see Appendix D table D1 and figure D1-D2) were considered in light of their interpretability and capability to effectively reduce the number of dimensions and capture the data's variance. Overall, FAMD does not perform well in either regard. To capture at least 50% of the data's variance at least 13 components. Compared to the original data, this is only a minor reduction of dimensions while at the same time losing half of the information of the data. Additionally, the variables' loadings on most components do not allow for a clear interpretation of their meaning. For a purely prediction-based task this would be less important. However, additionally to the prediction assessment, this study is interested in the contributions of the individual variables which would get strongly obscured using the FAMD components. It is for these reasons, that FAMD was not applied to the data in this study. Instead, machine learning techniques are selected with the focus on their ability to handle the issue of high dimensionality.

3.2. Machine Learning Models

To investigate which household characteristics, influence welfare and to what extent these factors can be utilised for accurate predictions, this study will conduct four different models and compare their results regarding prediction accuracy and feature importances. Specifically, it will be examined whether the strength and direction of the individual predictors are consistent across different models. Stable results across different model specifications could increase confidence in the accuracy of the model outputs which is key for effective policy development. Predictive accuracy and model performance will be assessed using the following metrics:

- Root Mean Squared Error (RMSE): Defined as the root of the sum of the squared differences between actual and predicted outcome. Due to the squaring, large residuals are penalised stronger. A lower value indicates better performance.

- Mean Absolute Error (MAE): Defined as the sum of the absolute differences between predicted and actual outcome values. The measure treats all residuals equally. A lower value indicates better performance.
- R-Squared: Defined as the proportion of the variance in the outcome variable which can be explained by the predictors included in the model. A higher value indicates better performance but also depends on the number of predictors in the model as well as the baseline variance in the outcome variable.

Model choices should always take a holistic view and consider multiple metrics which focus on different aspects of a model. These four indicators will provide a comprehensive understanding of the models' performances. Having defined the aim and focus of the analysis, the following machine learning models are conducted and will be elaborated in more depth in the following sections: Elastic Net, Random Forest, CatBoost, and SVM.

3.2.1. Elastic Net

The first model is an Elastic Net and belongs to the family of regularised regression techniques. It is commonly used for the purpose of variable selection and with the aim of obtaining a relatively simple, interpretable model. Elastic Net combines two penalties which are applied to the effect coefficients: the first penalty, L1, is akin to the penalty used in Lasso regression. It penalises coefficients by shrinking them to zero with the aim of reducing the influence of noise on effect estimates and increasing the reliability of model predictions. The second penalty, L2, is akin to that used in Ridge regression. Its aim is not to eliminate variables by shrinking their coefficient to zero but rather making them smaller and more stable, especially in the presence of multicollinearity.

Modelling the Elastic Net involves two key hyperparameters, namely the mixing parameter α and the parameter for regularisation strength λ. The α parameter ranges from zero to one. A value of zero indicates the exclusive application of the L2 penalty, while a value of one indicates the exclusive application of the L1 penalty. Values between zero and one represent a mixture of both penalties. The λ parameter can be zero or any larger, positive value. A value of zero effectively means no application of any penalties. A higher value results in stronger penalisation (i.e., either L1 or L2, depending on the chosen α), in extreme cases shrinking all variables down to zero. Both hyperparameters are tuned using 10-fold cross-validation as explained above.

Overall, the key advantages of this model are its ability to flexibly handle high-dimensional and multicollinear data while keeping interpretability high. Compared to a pure Lasso model which takes a stricter approach by reducing coefficients to zero rather than just shrinking them, there is potential of keeping some variables with low coefficients which are potentially adding noise rather than explanatory power. Depending on the task, this can be seen as an advantage or disadvantage. As this study is also somewhat exploratory in identifying variable associations with the outcome, the more lenient, flexible approach of the Elastic Net is positive. The key disadvantage of the Elastic Net is its assumption of linear relationships between the independent and dependent variables. Especially with complex data like these, this assumption is likely violated. Thus, while it offers a first opportunity to identify relevant variables and their influence on the outcome, other models are needed to explore and account for potential interactions and non-linearities between the variables. Tree-based methods are a common choice for this issue and will be explained next.

3.2.2. Tree-Based Models – Random Forest and Boosting

The basis of any tree-based model is a decision tree. It is a non-parametric modelling approach aiming to reduce a loss function (e.g., MSE) by sequentially splitting the data in a recursive, binary fashion using the available features. Specifically, at each split, also called node, the algorithm evaluates all potential splits across the data space and chooses the split that best reduces the loss function at this specific step, regardless of any potentially better splits in the future. Non-linearities and interactions can be easily modelled in this way because variables can be used multiple times in the process of constructing the tree. After all splits are completed, data points are assigned to their respective leaves, and the average outcome within each leaf is used as the prediction.

A further advantage of decision trees is its simple interpretation and visualisation. However, a major drawback of decision trees is their tendency to overfit as one decision tree simply represents the best and most detailed representation of the provided data. This leads to poor predictions on new, unseen data. Although there are methods to fight this issue (e.g., cost-sensitive pruning), their out-of-box performance often remains sub-optimal. There are two major adaptions to solve this issue which will be applied in this study, namely Bagging via a Random Forest and Boosting via the CatBoost Algorithm.

3.2.2.1. Bagging / Random Forest (RF)

Bagging and the construction of a Random Forest describes the process of randomly resampling a proportion of the data using bootstrapping and constructing a separate decision tree on each subsample. The final prediction of the training process is the average prediction across all fitted trees. Especially if there are a few strong predictor variables in the dataset which dominate the predictive power of other variables, the trees will be correlated with each other and the beneficial effect of the forest will be reduced. To decorrelate the trees, only a random subsample of all features is provided at each node. Depending on the implementation, there are several hyperparameters than can be adjusted to increase model performance and which in part function complementary to each other. Using cross-validation, the following parameters will be tuned: the maximum depth parameter influences the complexity of each individual tree. The minimum number of samples per leave regulates the number of further splits by setting a lower bound to the number of data points per leave. Similarly, the minimum number of samples required for a further split is tuned. Another parameter adjusts the number of variables that the algorithm can choose from to decide on the next split. Lastly, the number of trees grown can be varied.

3.2.2.2. Boosting

Boosting algorithms take a different approach and grow smaller, "weaker" trees sequentially. Fitted trees are not independent but are rather fitted on the previous trees' residuals. Thus, each new tree aims to reduce the errors of its predecessor to improve overall prediction accuracies. The number of sequential trees, their depth, and the overall learning rate which regulates the influence of each tree on the final prediction are hyperparameters that are tuned. There are several specific implementations of this algorithm including XGBoost, AdaBoost, and CatBoost. CatBoost can handle categorical data in an advanced, sophisticated way and will be used in this study as most of the independent variables are categorial. Specifically, CatBoost one-hot encodes only variables with at most two categories[1]. Variables with more categories are not one-hot encoded but rather transformed into numeric variables using ordered target statistics, thereby effectively avoiding the exponential increase in the number of dimensions due to dummy variables. Specifically, the values of the categories are replaced by their outcome-mean, though to avoid data leakage (i.e., the model "seeing" the

[1] Note that this is a parameter that could be adjusted. In this study, the default value was kept to transform as many variables as possible aiming to reduce dimensionality.

outcome that it should predict), the means are only calculated based on the previous rows (hence: *ordered* target statistics). Furthermore, each tree is fit on a random permutation of the original dataset leading to varied values for the ordered target statistics, further preventing data leakage.

While both Random Forest and CatBoost are state-of-the-art algorithms often performing well on complex data, the understanding and interpretation of the model beyond their predictions is more difficult as the results are a mix of many different samples, variables, and predictions. While relative feature importances can be calculated and give insight into the relative contribution of a variable in the overall predictions, effect sizes and directions are more obscured. However, advances in the realm of Explainable AI led to some tools that allow for a deeper understanding of the models' processes. One such tool (i.e., the SHAP library implemented in Python) will be explained in a later chapter and will be used to interpret the models' results.

3.2.3. Support Vector Regression (SVR)

SVR is an adaptation of the classic support vector machine originally developed for classification. In essence, SVR aims to find a function which deviates from the outcome values only by a pre-defined maximum margin. Deviations outside this margin are penalised. At the same time SVR aims to keep this function as simple as possible to avoid overfitting. SVR uses kernels which allows it to project the data into higher-dimensional spaces to model complex non-linearities. The following hyperparameters are tuned using cross-validation: first, the C parameter determines the regularisation strength. Epsilon indicates the maximum margin for error tolerance. Gamma regulates the kernel coefficient, and the kernel parameter indicates the kernel used. Common regression kernel specifications are the linear kernel, the radial basis function (rbf), and the polynomial kernel. While the flexible handling of the data in higher dimensional spaces is a powerful characteristic of SVR, the mathematical operations behind the algorithm make it more difficult to understand and interpret. Again, SHAP will be used to gain additional insights into the model results.

3.3. Shapley Additive Explanations (SHAP)

Accurate predictions are often not the only required output of machine learning models. Instead, understanding the models' decision-making process is of great importance when trying to develop new policies. Frequently, the only insight reported from machine learning models such as Random Forests, CatBoost, and SVMs are relative feature importances. While offering some information, these relative scores do not provide information on the direction or strength of the individual predictors. The SHAP package in Python is an Explainable AI tool aiming to improve this information deficit by making machine learning models more transparent and interpretable. The explanations offered here will provide a high-level understanding of SHAP and will provide explanations on how to interpret its output. The deeper mathematical processes underlying it are covered elsewhere in the literature (Molnar, 2025).

SHAP is based on a cooperative game theoretical background so that independent variables are conceptualised as the players, and the payout as the models' predictions. Essentially, SHAP calculates the contribution of each independent variable to a specific outcome (i.e., row in the training or test set) while ensuring that the difference between the average and predicted output is consistently distributed among all independent variables. A SHAP value can therefore be interpreted as the contribution of a respective variable to a specific prediction. The higher (and positive) the SHAP value, the more strongly the variable contributes to increasing the prediction, whereas the lower (and negative) the value, the more strongly the variable contributes to decreasing the models' prediction. Note that SHAP values are not standardised but are interpreted on the scale of the model's outcome variable, effectively enabling substantive interpretations of the predictions.

Figure 1 shows an example based on one data point extracted from the sample and its prediction is explained by one of the models. The base value indicates the overall sample's mean prediction. The bold number indicates the data point's predicted outcome. For each variable, one can see whether it increased (red) or decreased (blue) the base value to reach the predicted value. For example, the household's main information source were interpersonal sources, and this increased the predicted deviation from the advised scenario by 0.05 units. Figure 3 and 5 in the next chapter represent beeswarm plots that follow the same logic but present the findings for all data points, so that the general pattern becomes visible. This way, researchers can get a deeper understanding of the underlying workings rather than only aggregated, global feature importances.

Figure 1. SHAP example on one data point.

4. Results

The results section is divided into two parts. In the first part, the prediction accuracies of the four models are presented. In the second part, the predictor stability across models is evaluated using the regression output from the Elastic and the SHAP values for the Random Forest, the CatBoost, and the SVM model, respectively.

4.1. Prediction Accuracy

Table 1 shows the performance metrics of all models for both the cattle and the goat herders. Note, that cattle and goat herders can not be compared strictly as the models contain different number of rows and substantially differ in their purpose and economic value as explained above. Starting with the cattle dataset, the best performing model depends on the chosen metric. Based on the RMSE, CatBoost performs best with a score of 1.641. However, differences across all four models are minimal, with the largest gap being only 0.023 units (SVM, score: 1.664). A similar pattern can be observed when looking at the MAE. Here, the best performing model is the SVM with a score of 1.172, and the worst performing model the Elastic Net with a score of 1.189, resulting of a difference of only 0.03. Lastly, looking at the R-Squared the CatBoost model again performs best with a score of 1.5% explained variance. For this metric, the differences across models are slightly more pronounced. While the Random Forest explains only 0.7%, the Elastic Net explains only 1.0% of the variance in the outcome. Notably, the SVM performs worst with a negative R-Squared of -1.1%, indicating that the model's predictions are worse than simply predicting the mean. Still, the R-Squared is exceptionally low across all models, suggesting no substantial predictive power of the independent variables. That the CatBoost algorithm performs best in two of the three performance metrics emphasises its advantage of handling the categorical variables.

Moving on to the goat dataset, the SVM performs best across all metrics albeit differences across models are negligible. Specifically, the RMSE of the SVM lies at 1.363, but all other models' score is only 0.007 units higher (RMSE: 1.370). Similarly, there is only a difference 0.012 units, between SVM's MAE of 1.029 and the highest MAE of 1.045 (CatBoost). The SVM explains 2% of the variance in the outcome, the Random Forest's and Elastic Net's score is lowest at 1.0%, again showing only minor differences, and a generally low R-Squared across all models.

Table 1. Performance Metrics on Test Sets after Cross-Validation (Cattle & Goat)

		Cattle				Goat		
		RMSE	MAE	R2		RMSE	MAE	R2
Elastics Net		1.647	1.189	0.010		1.370	1.044	0.010
Random Forrest		1.648	1.185	0.007		1.370	1.041	0.010
CatBoost		**1.641**	1.187	**0.015**		1.370	1.045	0.011
SVM		1.664	**1.172**	-0.011		**1.363**	**1.029**	**0.020**
Note: Bold indicates the best values for a given metric – dataset combination.								

4.2. Predictor Stability

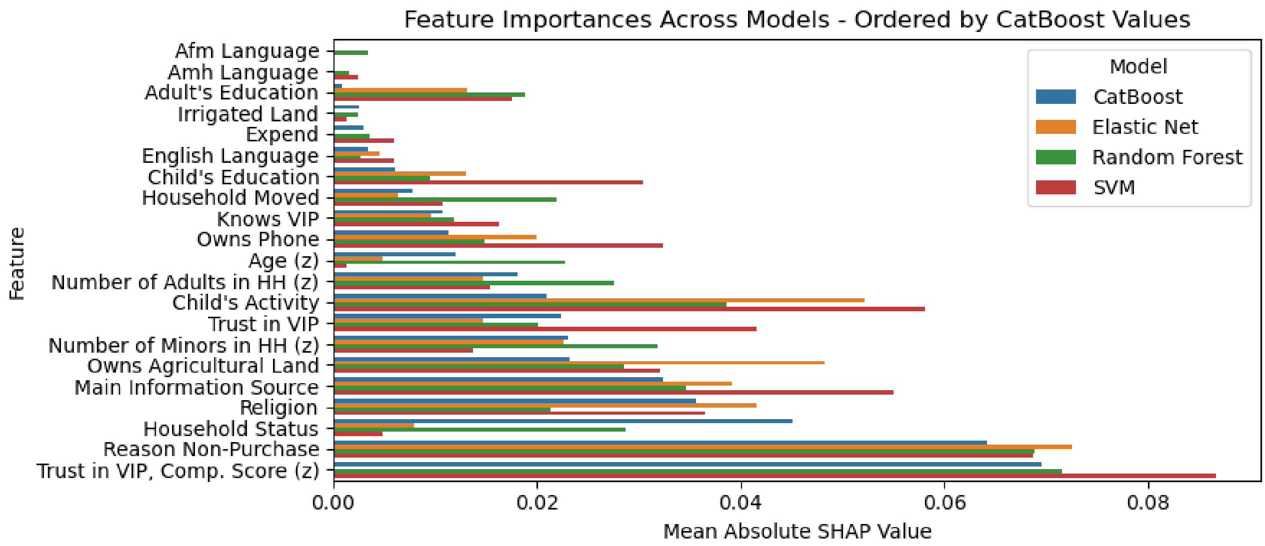
Figure 1 shows the mean absolute SHAP values across all four models for the cattle dataset. Values are ordered based on the overall best performing model (i.e., CatBoost).[2] The effects of the dummy variables in the Elastic Net, Random Forest, and SVM have been aggregated to their original categorical feature by calculating their absolute sum. In this way, they are both comparable to the CatBoost model which did not require dummy variables and represent the overall variables importance in the model. The figure shows that the effects across all models and effects are extremely low, with the strongest effect being merely 0.06 for the reason for non-purchase variable in the CatBoost and Elastic Net model. Therefore, note that while differences in effect sizes will be discussed, they are small. Note further, that all variable have been included in all models.[3]

Overall, the features' importances are stable across the different models, with only few deviations. The composite trust score, reason for non-purchase, and the main information source households used to get information about IBLI are consistently the strongest predictors of welfare outcomes. Agreement across models and relatively stronger effects can also be noted for the possession of agricultural land, the number of minor and adults in the household, religion, and possession of a phone. Interestingly, the child's and adult's education are a relatively weak predictor in the CatBoost model, but noticeably stronger in other models. Reversely, the household status is a relatively strong predictor for CatBoost and the Random Forest, but of much less importance for the SVM and Elastic Net. Language capabilities (i.e., Afan Omoro, Amharic, and English), as well as household expenditure habits, and owning land irrigation are generally some of the weakest predictors across models.

[2] Note that this is done mostly to ease interpretation and reading of the graph. As differences between model performances are overall small, another model could have been chosen to order the bars as well.
[3] The Elastic Net could have also been used as a baseline model to limit the number of variables in the subsequent models to only include those with a non-zero effect. This was not done because the study's aim was to investigate the behaviour of all variables across models and the models had their own methods to combat high dimensionality. Still, future research could base their variable selection on models like the Elastic Net.

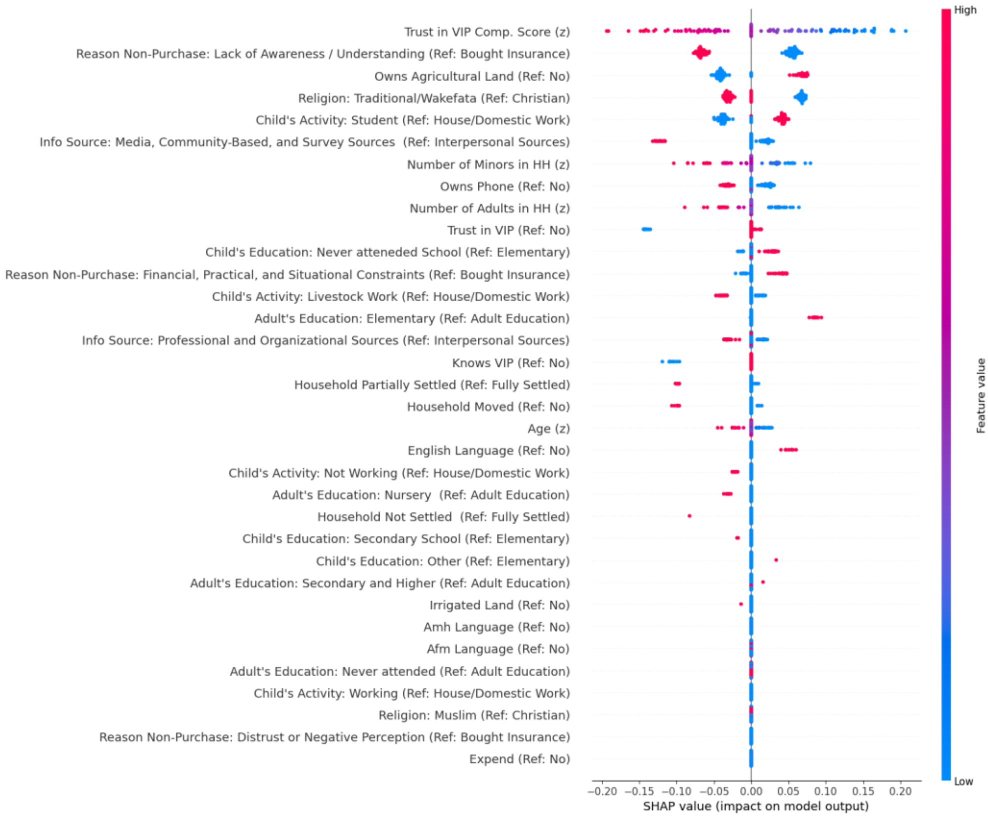
So far, only differences in absolute average effects (i.e., overall feature importances) have been discussed. However, the graph does not allow for any interpretation of the direction of the effects or the differences between categories of a variable. To get those deeper insights, a SHAP beeswarm graphs can be computed. Figure 2 show the results for the Random Forrest model, but the direction of effects and general conclusions are mostly similar across models (see Appendix F Figure F1-F3).[4] The effects for the strongest predictors will be briefly highlighted. The graph shows that households which did not purchase IBLI due to a lack of awareness or understanding compared to households which purchased IBLI, deviate less from the advised scenario and have therefore a higher estimated welfare. However, households which did not purchase IBLI due financial, practical, or situational constraints compared to household which purchased IBLI, show higher welfare deviations. More minors and adults in the household seems to decrease welfare deviations. Owning agricultural land is also associated with higher welfare deviations. Households which are either not or only partially settled also show lower welfare deviations compared to fully settled households. Further, households belonging to the traditional / Wakefata religion compared to the Christian religion show lower welfare deviations, but the effect is ambiguous for Muslims. Lastly, owning a phone compared to not owning a phone also yields lower welfare deviation predictions.

Figure 2. Feature Importances Across all Models – Cattle Herdes



[4] CatBoost cannot be used here, because of its treatment of categorical values. Thus, the Elastic Net is chosen as it is the second-best performing model.

Figure 3. Beeswarm Graph for Individual SHAP values of the Elastic Net– Cattle Herders



The next section presents the model results for the goat dataset. Figure 3 shows the mean absolute SHAP values across all four models, again with aggregated dummy variables and ordered based on the overall best performing model (i.e., SVM). Effects are generally small but largely consistent across models, similar to the previous results using the cattle data. Again, the composite trust score, reason for non-purchase and the main information source households used to get information on IBLI are the overall strongest predictors, followed by the household head's age and the number of minors in the household. The child's economic activity is a similarly strong predictor in all models but CatBoost, where it exerts practically no influence. Owning a phone has a noticeable influence in the Elastic Net and SVM models, but less so in the tree-based models. Reversely, the household status exerts a relatively strong influence on the welfare outcome in the tree-based methods, but less so in the Elastic Net and SVM model. Language capabilities, the usage of land irrigation, possession of agricultural land, knowing the VIP, children's education, the number of adults in the household, and religion are generally some of the weakest predictors across models. Thus, compared to the results of the cattle dataset, there are more features with little to no effect across all models. The outcome predictions are seemingly driven by fewer features, overall. Figure 4 shows the beeswarm graph for the SVM for deeper insights into the direction and distribution of the effects within the features. The results mirror those found on the cattle dataset. Reason for Non-Purchase, Main Information Source, and the composite trust score are consistently the strongest features. Appendix F Table F4-F6 shows the beeswarm graphs for the CatBoost, Elastic Net, and Random Forest.

Figure 4. Feature Importances Across all Models – Goat Herdes

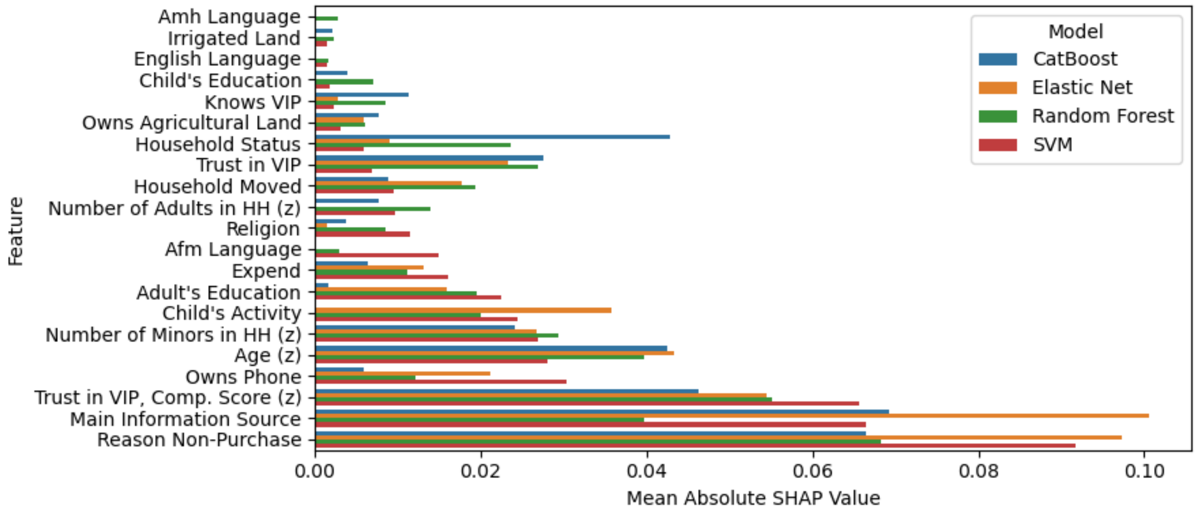


Figure 5. Beeswarm Graph for Individual SHAP values of the SVM – Goat Herders

5. Discussion and Suggestions for Improvement

The key insight from the above analyses regarding the reliability of household predictors in welfare estimation is that there is an overall high level of stability across models. The composite trust score, reason for not purchasing IBLI as well as the main information source of households were consistently the main predictors both for cattle and goat herders. Results show that not purchasing IBLI due to a lack of awareness or understanding is associated with lower deviations from the advised scenario / higher welfare than actually purchasing IBLI. Considering previous research indicating a tendency for excessive insurance uptake (Harrison et al., 2020), this finding may suggest a similar dynamic and supports the need for interventions that focus not merely on promoting IBLI uptake, but on enabling informed decision-making. Further, not purchasing IBLI due to financial, practical, and situational constraints (one category of the Reason for Non-Purchase variable) seems to decrease the households' welfare through larger deviations from the advised scenario. This finding emphasises the need to make IBLI accessible to households in need for protection from natural disasters.

Interpersonal contacts as the main information sources overall decreased welfare compared to other sources (i.e., Professional and Organizational Sources, Media, Community-Based, and Survey Sources), perhaps due to their potential to introduce personal biases into households' decision-making. This would support the need to expand efforts for wide-spread, professional information campaigns.

Higher trust in the IBLI insurance seems to increase welfare. However, the composite trust score contains some items that could be interpreted as cognitive understanding of IBLI insurance rather than representing trust. Thus, more research is needed to identify what this variable indicates to inform policy-making. While somewhat less stable in the relative strength of the effect across models, fully settled households showed lower welfare than both partially or not settled household in both the goat and cattle dataset. Previous research highlighted the importance of being able to move places cope with natural disasters (Taye, 2023). If fully settled households are more restricted in coping with shocks and make more welfare-reducing decisions in the IBLI context, this subgroup could benefit most from interventions.

The models included three indicators of different language skills, but neither substantively contributed to the predictions. From a data perspective, this could be because the variables show very little variation across categories (see Appendix B Table B1-B2) and therefore send little to no signal that can be used during model training. From a theoretical perspective, the weak influence of these variables can be seen as desirable as they could indicate little to no differences between social groups, especially if language differences are seen as proxies for ethnic group memberships.

Although the predictors are generally consistent across models and datasets, their effects are vanishingly small, and results show extremely weak prediction accuracy with predictions centring around the outcomes' means. Multiple reasons might explain this finding. First, the high number of categorical features which are mostly one-hot encoded lead to reduced variation in the data space making it sparser and more difficult to exploit variation. Additionally, there is low spread across the categories of a given categorical feature, exacerbating the problem of low variance (see Table B1 in the Appendix). To explore whether using more continuous features could increase the models' performances, the models were rerun using the adult's and the child's education variable as a numeric indicator. This was not done originally, as the ordinal ordering of certain categories (e.g., the category "adult education") was difficult to establish. The results are reported in Appendix G Table G1[5]. Overall, the performance of the R-Squared is slightly higher (around 0.1 and 0.2 percentage points) for most models. This is especially interesting since the variables did not exert strong influence in the previous models, and some dummy variables were even eliminated by the Elastic Net. As the model still contains many categorical features which cannot be transformed numerically and this has just been a small change to the model, the adjustment can be seen as promising. Future research which might have access to more features should further explore this option and aim to use continuous features where possible.

Another way to potentially improve model performance is to focus on predicting specific strata of the outcome variable, rather than simply its mean, which is typically the default. By focusing on different quantiles instead of the mean or median, one might uncover features which relate differently across the outcome distribution. In such a case, a global model might not be the most appropriate choice for modelling welfare outcomes. Instead, this would encourage a separate investigation of low or high welfare outcomes, both empirically and theoretically. While models like Random Forest, CatBoost, and SVM can capture complex interactions and are therefore capable of modelling different parts of the distribution, quantile regression offers a simple and intuitive way to explore these relationships more visually. For explorative purposes, quantile regression was applied to the cattle dataset. Results are reported in Appendix H Figure H1-H4. The results show that the focus on specific quantiles of the outcome distribution rather than its mean neither improves the predictions nor shows variations

[5] Note that due to time constraints, the results reported in this chapter do not include the new composite trust variable. Nonetheless, the conclusions remain the same.

among the features. Still, with more continuous variables and more variation among the features and their categories, the approach could yield more insights and different results and is therefore worth further exploring in future research.

6. Conclusion

This study aimed to investigate which household characteristics predict welfare from IBLI insurance, to what extent state-of-the-art machine learning techniques can improve the predictive accuracy of welfare outcome estimations using these characteristics, and to what extent the identified characteristics are reliable across models? For this purpose, data from an IBLI insurance product marketed in Southern Ethiopia has been employed, and various machine learning models (i.e., Elastic Net, Random Forest, CatBoost, SVM) have been trained, examined and compared. Previous research in this field mostly focused on insurance uptake but neglected more specific welfare outcomes which allow for variation in the decision qualities, and often reported contradictory findings (Jensen et al., 2018; Takahashi et al., 2019).

The results in this study show that the included household characteristics do not explain any substantial variation in welfare outcomes in line with Jensen et al. (2018). Across all models, the test predictions centre around the mean. The little variance that is explained (between 0.2 and 1.2%), however, shows high stability in the feature importances and the direction of the effects. This indicates that while prediction accuracy is extremely low, the explained variance is not (only) due to noise but shows some signs of systematic patterns. More specifically, in all four models the composite trust score, the non-purchase reason and the main IBLI information source were the two strongest predictors. Not buying IBLI due to low understanding or awareness seemed to protect households from excess purchase, while not purchasing IBLI due to financial, practical, or situational constraints resulted in larger welfare losses. Albeit, low in effect sizes, this finding lends some support to the importance of reducing barriers for IBLI insurance in the region as well as increasing knowledge and awareness of the product. Further, receiving IBLI information from interpersonal contacts rather than organisational or medial sources seems to increase welfare losses. In combination with the previous finding, this emphasises the need for high-quality information campaigns. Lastly, there is some indication that fully settled household suffer larger welfare losses. Due to their potentially increased vulnerability to the consequences of natural shocks (Taye, 2023), this group of households could be seen as an important target group.

Despite these insights, the study has several limitations which future research could improve. First, many of the included variables show little variation making it hard to extract meaningful patterns. This is exacerbated by the large number of categorical and low number of continuous variables. Future research should aim to capture features with larger variance, as some exploratory analyses already showed some improvements. Further, the study mostly included characteristics that are largely unassociated with the households' livestock assets as those variables (e.g., size of the livestock, ratio of livestock covered by the purchases insurance) were used to construct the outcome. Still, this information is critical when examining the actual welfare gains and losses as they are relative to the household endowments. A ratio variable can be helpful to combat this issue but could only be exploratorily investigated in this study. Importantly, the distribution of the ratio variable shows that the vast majority of households already fulfil their welfare potential to 90% or more, potentially indicating a ceiling effect which makes it difficult to find factors that close the small remaining gaps. Future research is advised to explore this further. Lastly, the findings regarding the predictors should be considered with great caution as their effects are extremely low and do not allow for any causal claims. Future research should reconsider the variables included in welfare estimation and critically think of their relevance and measurement. I advise to focus and improve on this aspect rather than trying ever more complex models. Neural networks, for example, might seem appealing but typically only work well with a much larger dataset. In this context, increased model complexity most likely does not mend the issues introduced by data quality.

Nonetheless, this study contributed new insights to the literature: consumer and household characteristics are not enough to accurately predict welfare, the scope of variables most be extended to identify drivers of welfare from IBLI insurance, for example including more cognitive variables. Trust, non-purchase motivations and barriers as well as information sources might play a role in IBLI welfare but need further investigation. This information can be used by researchers to further the understanding of welfare from IBLI products. With more research building on these findings, policy and product design can be informed in the future.

Literature

Athey, S., Keleher, N., & Spiess, J. (2025). Machine learning who to nudge: causal vs predictive targeting in a field experiment on student financial aid renewal. *Journal of Econometrics*, 105945.

Athey, S., Simon, L. K., Skans, O. N., Vikstrom, J., & Yakymovych, Y. (2023). *The heterogeneous earnings impact of job loss across workers, establishments, and markets*. arXiv preprint arXiv:2307.06684.

Barrett, C. B., Harrison, G. W., Jensen, N., Morsink, K., Schneider, M., Swarthout, J. T., & Upton, J. (2024). Do No Harm: Evaluating the Welfare Effects of Behavioral Insurance *Interventions in Ethiopia*.

Bertram-Huemmer, V., & Kraehnert, K. (2017). Does Index Insurance Help Households Recover from Disaster? Evidence from IBLI Mongolia. *American Journal of Agricultural Economics*, 100(1), 145–171. https://doi.org/10.1093/ajae/aax069

Chantarat, S., Mude, A. G., Barrett, C. B., & Turvey, C. G. (2017). Welfare impacts of index insurance in the presence of a poverty trap. *World Development*, 94, 119-138.

Chantarat, S., Mude, A. G., Barrett, C. B., & Carter, M. R. (2013). Designing Index-Based livestock insurance for managing asset risk in Northern Kenya. *Journal of Risk & Insurance*, *80*(1), 205–237. https://doi.org/10.1111/j.1539-6975.2012.01463.x

Dercon, S. (1998). Wealth, risk and activity choice: cattle in Western Tanzania. *Journal of Development Economics*, *55*(1), 1–42. https://doi.org/10.1016/s0304-3878(97)00054-0

Gehring, K., & Schaudt, P. (2024). Insuring Peace: Index-Based Livestock Insurance, Droughts, and Conflict. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4702292

Harrison, G. W., Morsink, K., & Schneider, M. (2020). *Do no harm? The welfare consequences of behavioural interventions* (Vol. 12). CEAR Working Paper 2020.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.

Handel, B. R., Kolstad, J. T., Minten, T., & Spinnewijn, J. (2020). *The social determinants of choice quality: evidence from health insurance in the Netherlands* (No. w27785). National Bureau of Economic Research.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer.

Jensen, N. D., Mude, A. G., & Barrett, C. B. (2018). How basis risk and spatiotemporal adverse selection influence demand for index insurance: Evidence from northern Kenya. *Food Policy*, 74, 172–198. https://doi.org/10.1016/j.foodpol.2018.01.002

Kahiu, N., Anchang, J., Alulu, V., Fava, F. P., Jensen, N., & Hanan, N. P. (2024). Leveraging browse and grazing forage estimates to optimize index-based livestock insurance. *Scientific Reports*, 14(1). https://doi.org/10.1038/s41598-024-62893-4

Kaumbata, W., Banda, L., Mészáros, G., Gondwe, T., Woodward-Greene, M., Rosen, B. D., Van Tassell, C. P., Sölkner, J., & Wurzinger, M. (2020). Tangible and intangible benefits of local goats rearing in smallholder farms in Malawi. *Small Ruminant Research*, 187, 106095. https://doi.org/10.1016/j.smallrumres.2020.106095

Molnar, C. (2025). *Interpreting Machine Learning Models with SHAP*. Retrieved June 9, 2025, from https://christophm.github.io/interpretable-ml-book/shapley.html

Taye, M. (2023). Livestock insurance in southern Ethiopia: Calculating risks, responding to uncertainties. In I. Scoones (Ed.), *Pastoralism, uncertainty and development* (pp. 93-105). Practical Action Publishing. https://practicalactionpublishing.com/book/2667/pastoralism-uncertainty-and-development

Tafere, K., Barrett, C. B., & Lentz, E. (2018). Insuring Well-Being? Buyer's remorse and peace of mind effects from insurance. *American Journal of Agricultural Economics*, *101*(3), 627–650. https://doi.org/10.1093/ajae/aay087

Takahashi, K., Ikegami, M., Sheahan, M., & Barrett, C. B. (2015). Experimental evidence on the drivers of Index-Based livestock insurance demand in Southern Ethiopia. *World Development*, 78, 324–340. https://doi.org/10.1016/j.worlddev.2015.10.039

Takahashi K., Noritomo Y., Ikegami M., & Jensen N. D. (2019, October 23). *Understanding Pastoralists' Dynamic Insurance Uptake Decisions: Evidence from Four-year Panel Data in Ethiopia*. Copyright © 2007-2008. https://grips.repo.nii.ac.jp/records/1723

UN Office for the Coordination of Humanitarian Affairs. (2024, January 10). *Ethiopia: Situation report* (Situation Report No. 10). Retrieved June 9, 2025, from https://www.unocha.org/publications/report/ethiopia/ethiopia-situation-report-10-jan-2024

Zewdie, Y., Taye, M., & Fava, F. (2020, July). Livestock insurance for pastoralists in Ethiopia: Exploring opportunities for scaling (ILRI Policy Brief No. 29). *International Livestock Research Institute*. https://cgspace.cgiar.org/handle/10568/110918

Appendix A

This section documents how original categories were collapsed to create the categorical variables used in this study. In the left column of the tables, you can see the category used in this study. In the right column are the original categories which were collapsed.

Table A1. Children's Education

Category in this study	*Original categories which were collapsed*
Never attended school	Never attended school
Elementary	Nursery, Grade 1 to Grade 8
Secondary	Secondary 9 – Secondary 12, Diploma
Other	Alternative Basic Education (ABE), Adult Education, Degree, Postgraduate, Other, specify, Technical education, Certificate level teachers

Table A2. Children's Main Activity

Category in this study	*Original categories which were collapsed*
Student	Student
Not working	Not working: too young, Looking for job, Not working: unable, Not working: too old
Working with Livestock	Herding household-owned livestock, Livestock production e.g. milking, sale of livestock products
House/domestic work	House/domestic work, Unpaid work in family's shop/business
Working	Petty trading e.g charcoal/water trading, Casual labor e.g. herding for pay, Livestock trading/broker, Shop/business owner, Wage/salaried employment, Mining, Farming non-livestock, Other, specify

Table A3. Adult's Education

Category in this study	*Original categories which were collapsed*
Never attended school	Never attended school
Nursery	Nursery
Elementary	Grade 1 to Grade 8
Secondary and Higher	Secondary 9, Secondary 10, Preparatory 11, Preparatory 12, Diploma, Degree, Postgraduate, Technical education
Adult Education	Alternative Basic Education, Adult Education

Table A4. Main Information Source

Category in this study	*Original categories which were collapsed*
Interpersonal Sources	Enumerators, a Kebele manager, NGO staff. development agents (DAs) or other government officials, neighbors, friends and relatives in informal groups
Professional and Organizational Sources	insurance extension staff (including VIPs/Insurance promoters), Oromia Insurance Company staff (OIC) excluding VIPs/insurance promoters and/or Oromia Saving and Credits Share Company (WALQO)
Media, Community-Based, and Survey Sources	network groups, the radio, community meetings, survey conducted by ILRI, posters, cooperatives, discount coupons, the television (TV)

Table A5. Reason for non-Purchase

Category in this study	*Original categories which were collapsed*
Bought Insurance	Bought Insurance
Lack of Awareness or Understanding	Did not understand insurance well enough, Did not understand coupon, Unaware/Have not heard that insurance was available, Did not know who the agent was or could not find them
Distrust or Negative Perception	Afraid of uncertainty in insurance, Did not trust the insurance agent or company, Discouraged by someone in the community/family, Did not want product/product not good for me, Did not want product/product not good generally
Financial, Practical, and Situational Constraints	Cost is too high, Too expensive/could not afford, Coupon got lost, Can rely on family and friends, Little/No risk of drought this year Other, Too busy, Waiting to see what happens to other people with insurance, Had coverage from KLIP, Could not because did not have ID and/or Mpesa account, Did not have enough animals

Table A6. Religion

Category in this study	*Original categories which were collapsed*
Traditional/Wakefata	Traditional/Wakefata
Muslim	Muslim
Christian	Prtotestant, Orthodox, Catholic, Other Christian

Appendix B

The tables below show summary statistics separately for the cattle and goat dataset.

Table B1. Summary Statistics for the Cattle Dataset

Variable	Categories	Range	Mean.	Std Dev
Welfare – Difference (before transformation)	-	-27,730.31–0.0	-723,68	1146,75
Age	-	14–96	43,24	16,61
Number of Minors in HH	-	0–9	3,26	1,63
Number of Adults in HH	-	0–18	2,28	1,33
Afan Omoro Language	No	-	0,81	-
	Yes	-	0,19	-
Agricultural Land	No	-	0,6	-
	Yes	-	0,4	-
Amharic Language	No	-	0,93	-
	Yes	-	0,07	-
Highest Education of Adult	Never attended	-	0,68	-
	Elementary	-	0,13	-
	Nursery	-	0,11	-
	Adult Education	-	0,05	-
	Secondary and Higher	-	0,04	-
English Language	No	-	0,92	-
	Yes	-	0,08	-
HH Expends More Than Median	Yes	-	0,25	-
	No	-	0.75	-
Irrigated Land	Yes	-	0,09	-
	No	-	0.91	-
Education of Child	Elementary	-	0,5	-
	Never attended school	-	0,43	-
	Secondary	-	0,06	-
	Other	-	0,01	-
Main Activity of Child	Student	-	0,52	-

	Working with Livestock	-	0,23	-
	Not working	-	0,18	-
	House/domestic work	-	0,06	-
	Working	-	0,01	-
Household Description	Fully settled	-	0,9	-
	Partially settled	-	0,09	-
	Not settled	-	0,01	-
Main Information Source	Interpersonal Sources	-	0,57	-
	Professional and Organizational Sources	-	0,3	-
	Media, Community-Based, and Survey Sources	-	0,13	-
Religion	Traditional/Wakefata	-	0,65	-
	Muslim	-	0,23	-
	Christian	-	0,12	-
Owns Phone	Yes	-	0,37	-
	No	-	0.63	-
Household Moved Last 6 Months	No	-	0,91	-
	Yes	-	0,09	-
Reason for Non-Purchase	Lack of Awareness or Understanding	-	0,42	-
	Financial, Practical, and Situational Constraints	-	0,31	-
	BOUGHT INSURANCE	-	0,2	-
	Distrust or Negative Perception	-	0,07	-
Know Vip	Yes	-	0,94	-
	No	-	0,06	-
Trust in Vip	Yes	-	0,92	-
	No	-	0,08	-

Table B2. Summary Statistics for the Goat Dataset

Variable	Categories	Range	Mean	Std Dev
Welfare - Difference	-	-10719.69–0.0	-286,86	463,76
Age	-	14–97	43,19	16,77
Number of Minors in HH	-	0–9	3,23	1,63
Number of Adults in HH	-	0–18	2,25	1,29

Afan Omoro Language	No	-	0,81	-
	Yes	-	0,19	-
Agricultural Land	No	-	0,6	-
	Yes	-	0,4	-
Amharic Language	No	-	0,93	-
	Yes	-	0,07	-
Highest Education of Adult	Never attended	-	0,68	-
	Elementary	-	0,13	-
	Nursery	-	0,11	-
	Adult Education	-	0,04	-
	Secondary and Higher	-	0,03	-
English Language	No	-	0,91	-
	Yes	-	0,09	-
HH Expends More Than Median	Yes	-	0,25	-
	No	-	0.75	-
Irrigated Land	Yes	-	0,09	-
	No	-	0.91	-
Education of Child	Elementary	-	0,51	-
	Never attended school	-	0,42	-
	Secondary	-	0,05	-
	Other	-	0,01	-
Main Activity of Child	Student	-	0,53	-
	Working with Livestock	-	0,22	-
	Not working	-	0,18	-
	House/domestic work	-	0,06	-
	Working	-	0,01	-
Household Description	Fully settled	-	0,91	-
	Partially settled	-	0,08	-
	Not settled	-	0,01	-
Main Information Source	Interpersonal Sources	-	0,6	-
	Professional and Organizational Sources	-	0,28	-
	Media, Community-Based, and Survey Sources	-	0,12	-
Religion	Traditional/Wakefata	-	0,65	-
	Muslim	-	0,24	-
	Christian	-	0,11	-
Owns Phone	Yes	-	0,36	-
	No	-	0.64	-
Household Moved Last 6 Months	No	-	0,91	-
	Yes	-	0,09	-

Reason for Non-Purchase	Lack of Awareness or Understanding	-	0,4	-
	Financial, Practical, and Situational Constraints	-	0,32	-
	BOUGHT INSURANCE	-	0,21	-
	Distrust or Negative Perception	-	0,06	-
Know Vip	Yes	-	0,94	-
	No	-	0,06	-
Trust in Vip	Yes	-	0,93	-
	No	-	0,07	-

Appendix C

The graphs below show the distribution of the outcome variable before and after the log+1 transformation. Note that the +1 was necessary to keep the zero values.

Figure C1. Distribution of Outcome Variable Before Transformation – Cattle Dataset

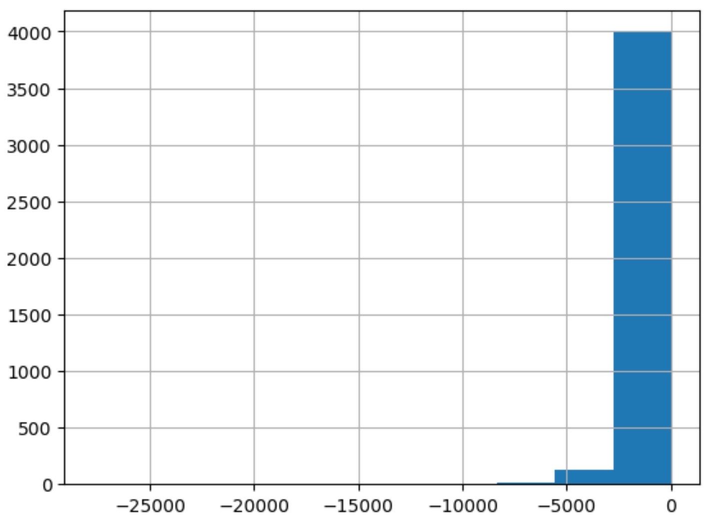


Figure C2. Distribution of Outcome Variable After Log+1 Transformation – Cattle Dataset

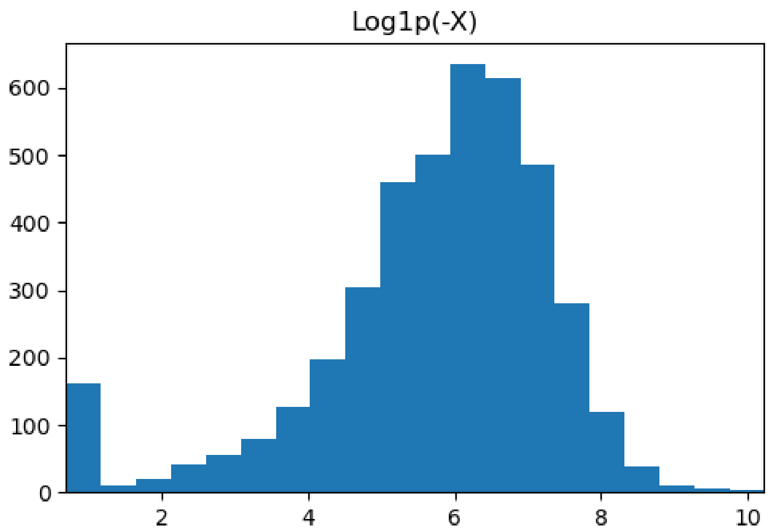


Figure C3. Distribution of Outcome Variable Before Transformation – Goat Dataset

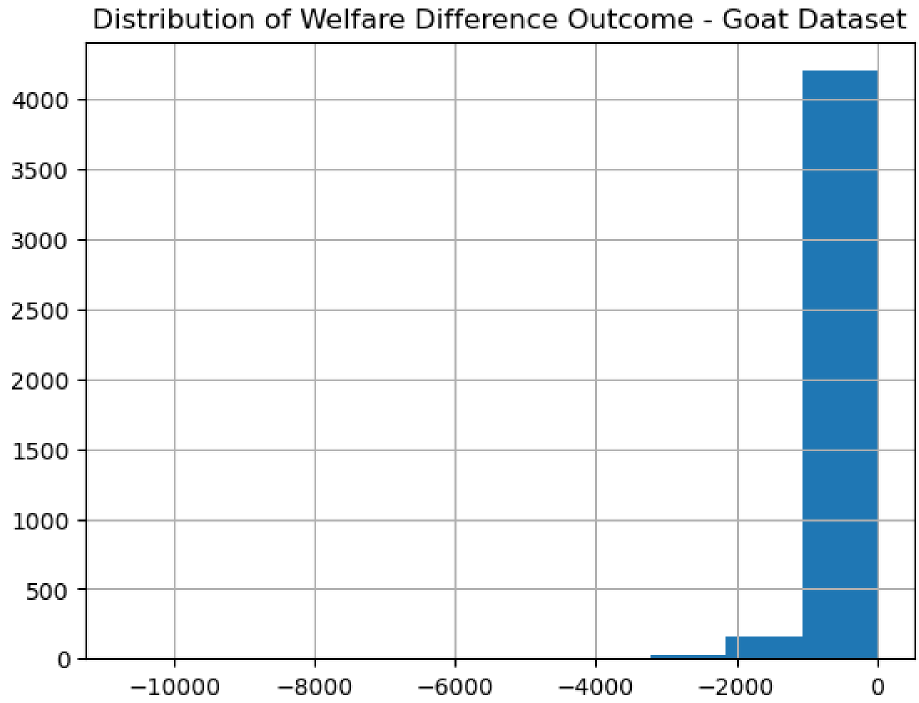
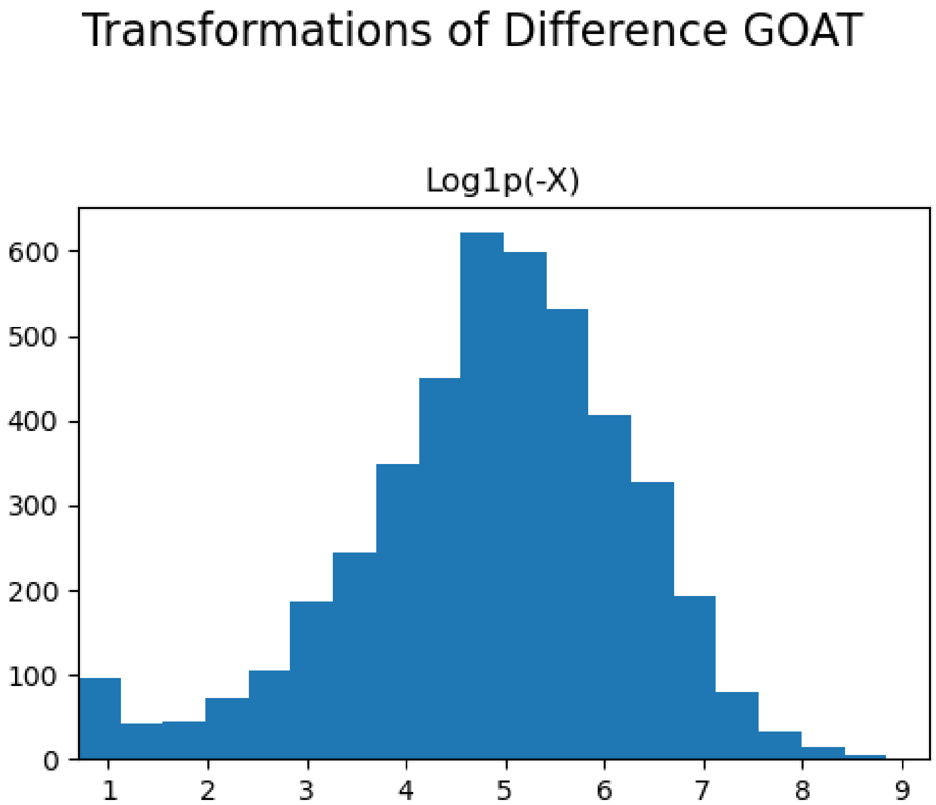


Figure C4. Distribution of Outcome Variable After Log+1 Transformation – Goat Dataset



Appendix D

This section documents the results of the FAMD analyses based on the Cattle Dataset.

Table D1. Eigenvalue, Variance (%), and Cumulative Variance (%) for a 20-Component FAMD solution.

Dimension	Eigenvalue	Variance (%)	Cumulative Variance (%)
1	2,33	6,67	6,67
2	2,07	5,91	12,58
3	1,68	4,81	17,39
4	1,50	4,30	21,69
5	1,39	3,98	25,67
6	1,37	3,91	29,58
7	1,28	3,67	33,25
8	1,14	3,26	36,51
9	1,11	3,17	39,67
10	1,09	3,12	42,79
11	1,66	3,03	45,82
12	1,06	3,02	48,84
13	1,05	2,99	51,84
14	1,03	2,95	54,79
15	1,01	2,88	57,68
16	1,00	2,85	60,53
17	0,99	2,82	63,34
18	0,98	2,80	66,14
19	0,95	2,72	68,86
20	0,94	2,70	71,56

Figure D1. Scree Plot visualising the Distribution of Eigenvalues Across Components

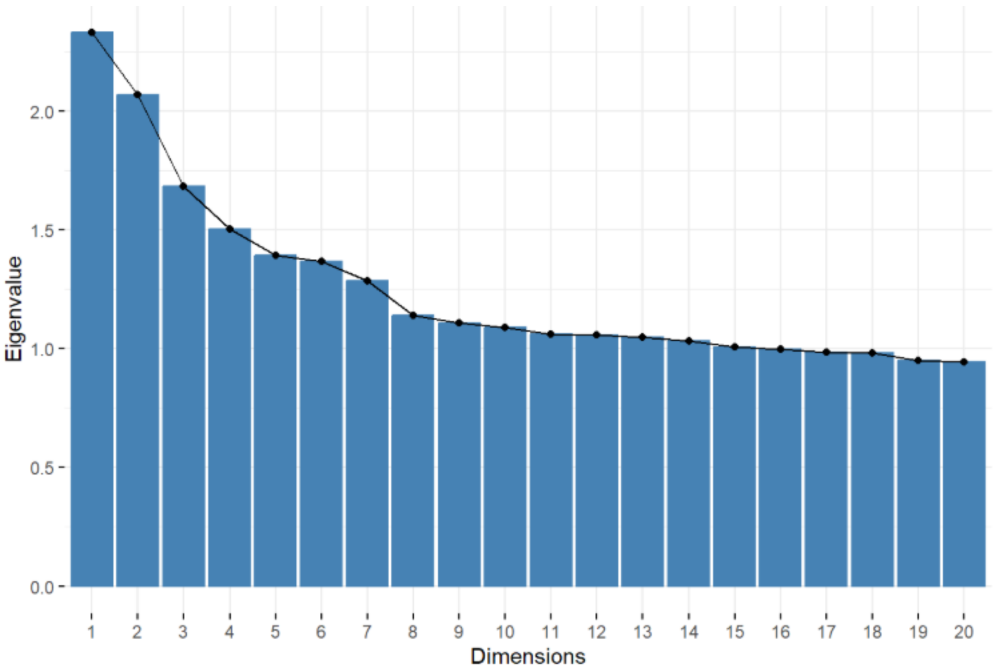
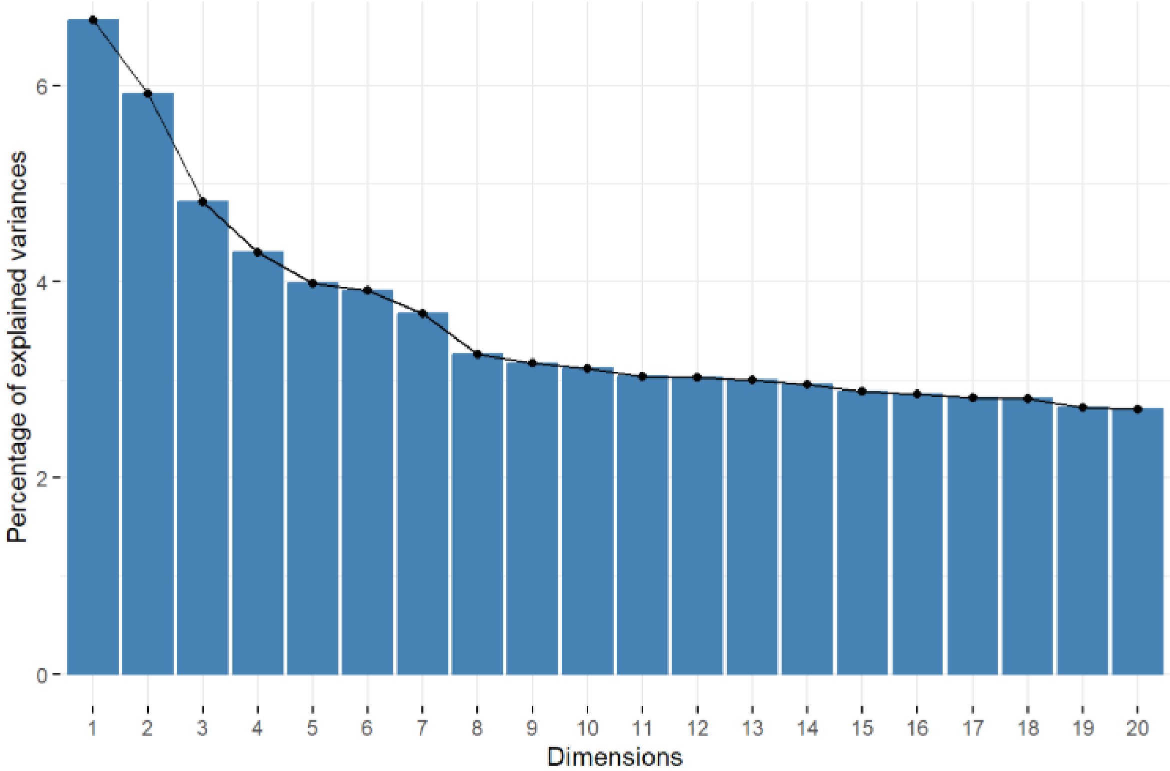


Figure D2. Scree Plot visualising the Distribution of Explain Variance Across Components



Appendix E

This section reports the results of the random effects models from which the residuals were extracted and used as the outcome variables.

Figure E1. Model Results for the Random Effect Model – Cattle Dataset

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: cs_diff_log ~ 1 + (1 | id)
##    Data: cattle_s
##
## REML criterion at convergence: 16254.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3353 -0.3742  0.1793  0.6311  2.5210
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 0.1416   0.3763
##  Residual             2.8230   1.6802
## Number of obs: 4142, groups:  id, 2189
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  5.69325    0.02747   207.2
```

Intraclass Correlation Coefficient (ICC) Cattle

```
# compute icc
icc <- 0.1416 / (0.1416 + 2.8230) *100
icc
```

```
## [1] 4.776361
```

Figure E2. Original Log+1 Transformed Variable vs. Extracted Residuals – Cattle Dataset

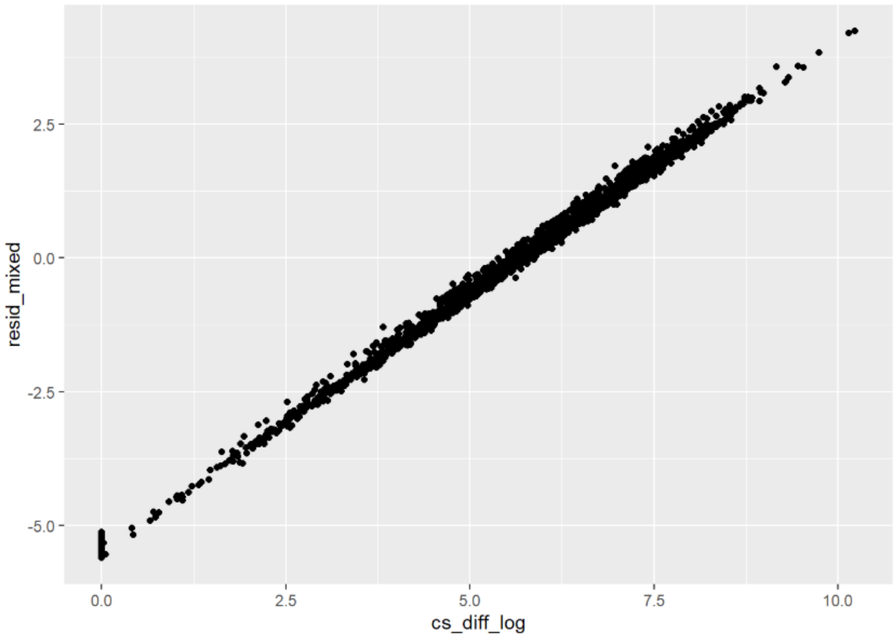


Figure E3. Model Results for the Random Effect Model – Goat Dataset

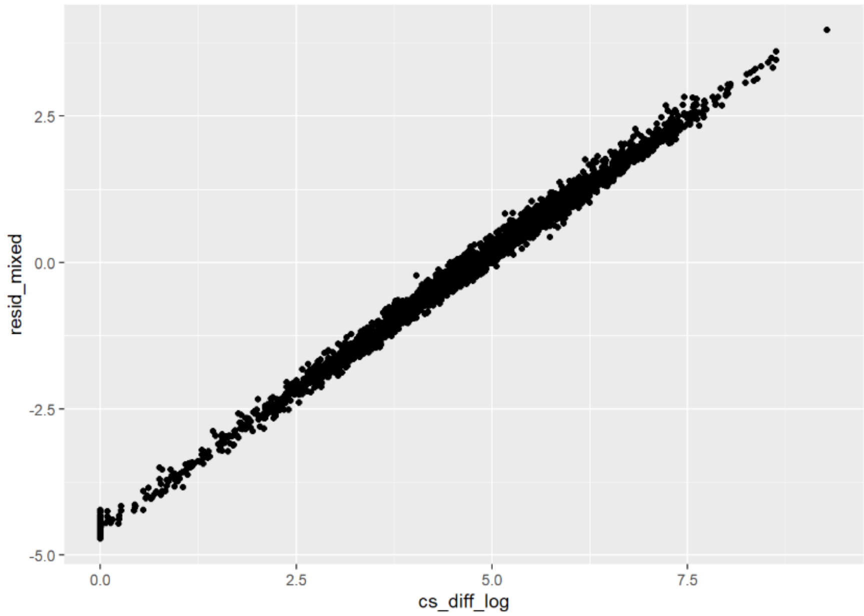
```
## Linear mixed model fit by REML ['lmerMod']
## Formula: cs_diff_log ~ 1 + (1 | id)
##    Data: goat_s
##
## REML criterion at convergence: 15938.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3027 -0.4933  0.0991  0.6389  2.7768
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 0.1459   0.382
##  Residual             2.0435   1.430
## Number of obs: 4403, groups:  id, 2235
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  4.82938    0.02316   208.6
```

Intraclass Correlation Coefficient (ICC) Cattle

```
# compute icc
icc <- 0.1459 / (0.1416 + 2.0435) *100
icc
```

```
## [1] 6.67704
```

Figure E4. Original Log+1 Transformed Variable vs. Extracted Residuals – Goat Dataset



Appendix F

This section reports the complete beeswarm plots for the remaining models not covered in the main text. That is the Elastic Net, SVM, and Random Forest for the cattle dataset, and the Elastic Net, Random Forest, and CatBoost for the goat dataset.

Figure F1. Beeswarm Plot for the Elastic Net – Cattle Dataset

Figure F2. Beeswarm Plot for the Random Forest – Cattle Dataset

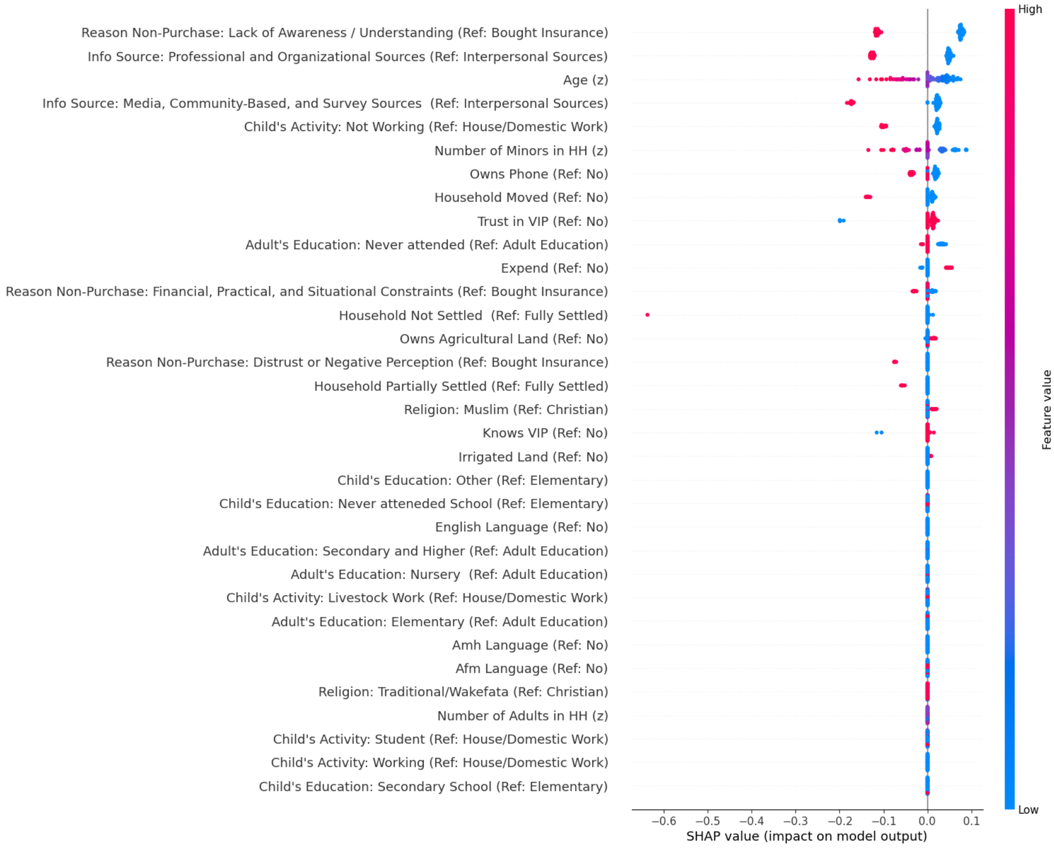


Figure F3. Beeswarm Plot for the SVM – Cattle Dataset

Figure F4. Beeswarm Plot for the Elastic Net – Goat Dataset

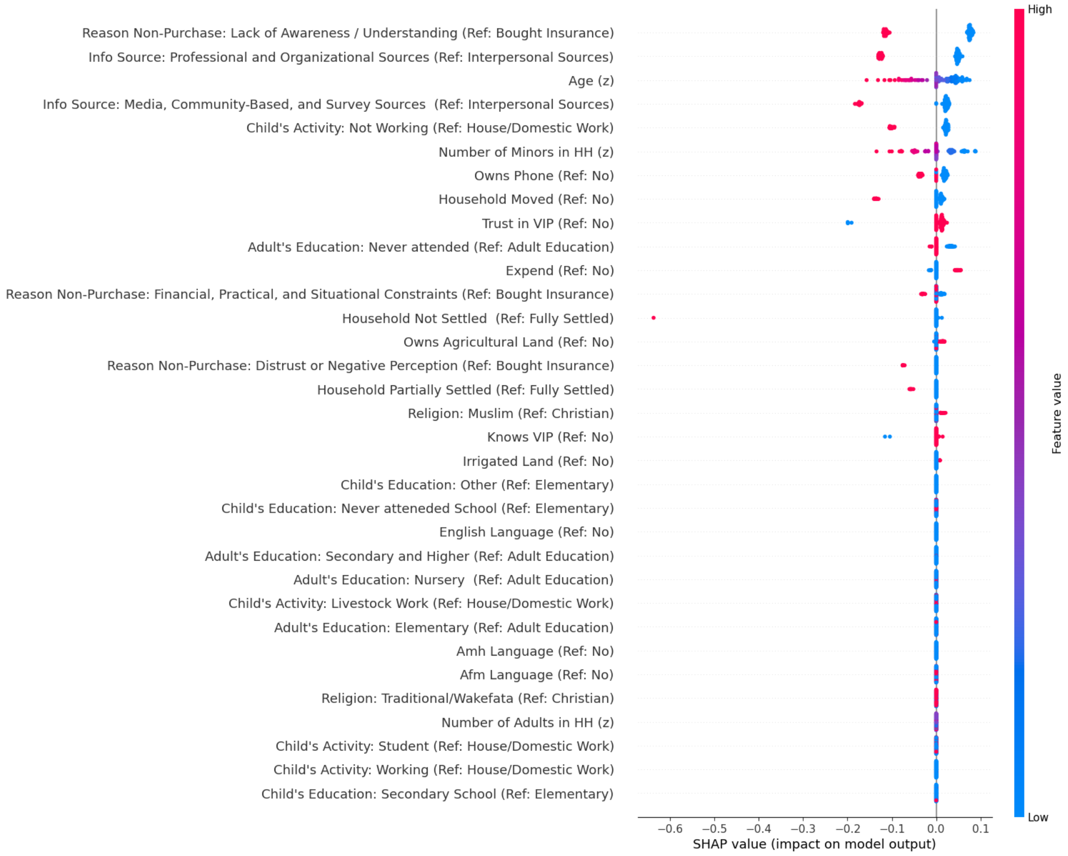


Figure F5. Beeswarm Plot for the Random Forest – Goat Dataset

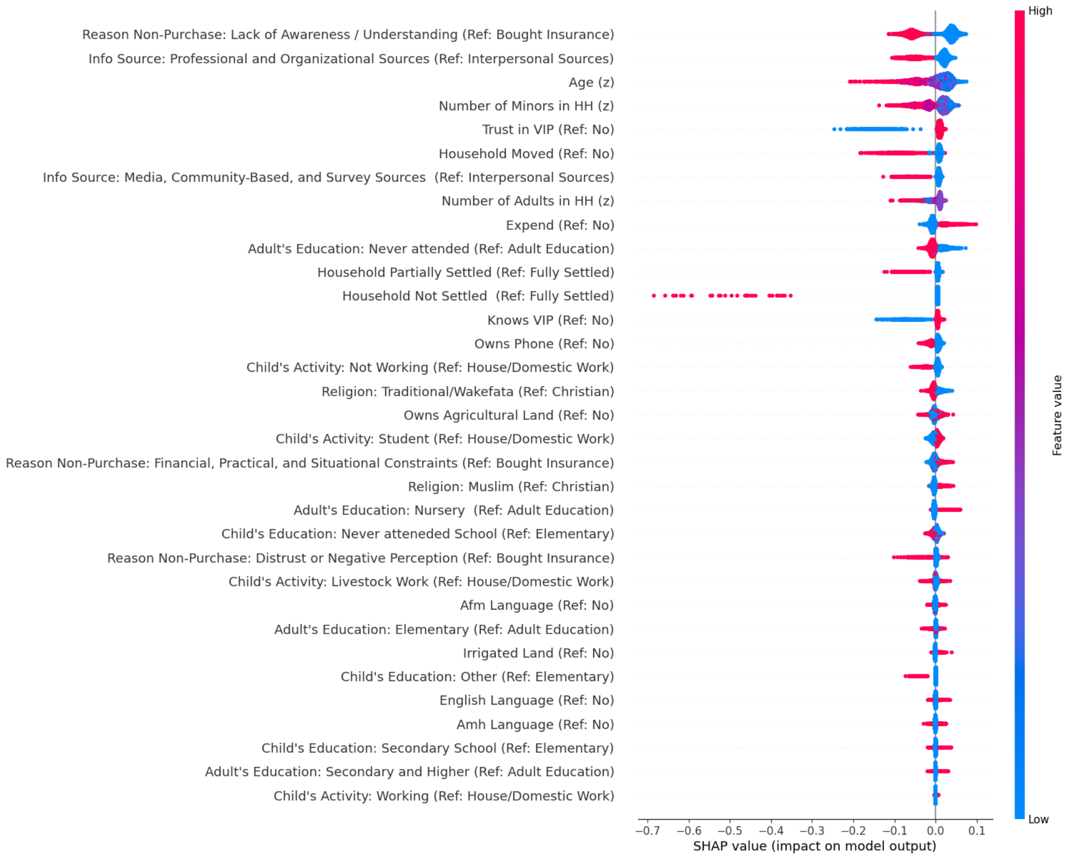
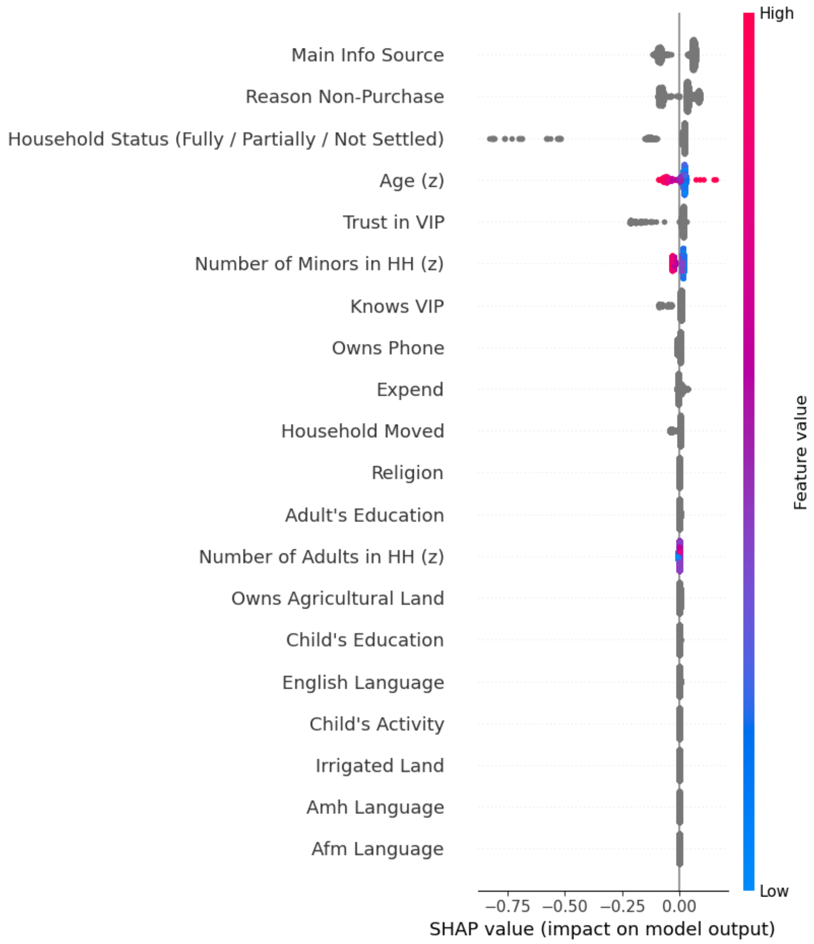


Figure F6. Beeswarm Plot for the CatBoost – Goat Dataset



Appendix G

This section reports the performance metris for the models with continuous educational variables (i.e., children's and adults' education).

Table G1. Performance Metrics for Models with Numeric Education Variables after (Cattle & Goat Datasets)

		Cattle				Goat		
		RMSE	MAE	R2		RMSE	MAE	R2
Elastics Net		1.650	1.190	0.005		1.371	1.041	0.009
Random Forrest		1.648	1.185	0.007		1.375	1.045	0.002
CatBoost		1.641	1.187	0.015		1.370	1.044	0.009
SVM		1.664	1.172	-0.011		1.369	1.029	0.011

Appendix H

This section reports the results from the Quantile Regression and Quantile Boosting (using CatBoost) which were conducted to explore potential model improvements. Only the cattle dataset has been used.

Figure H1. Linear Quantile Regression Coefficients Across Quantiles

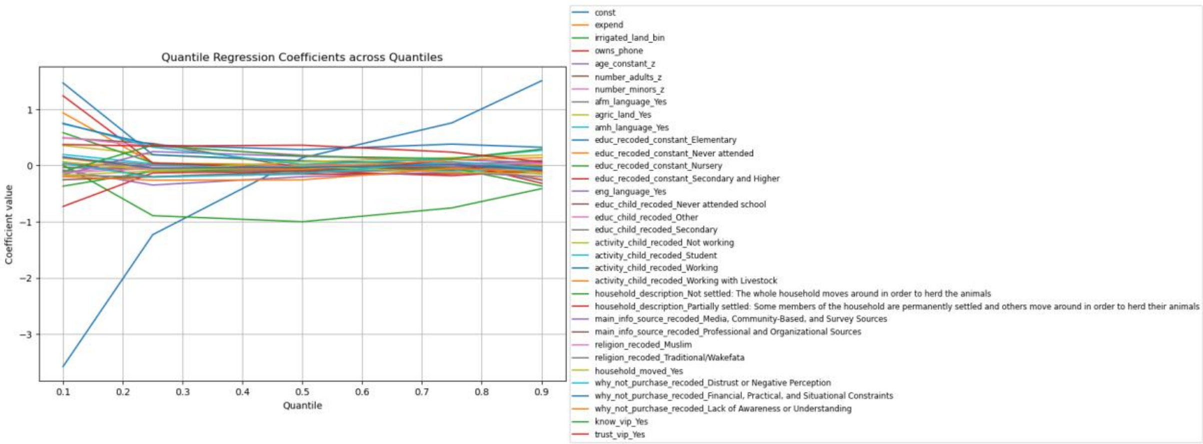


Figure H2. Predicted vs. Actual Outcome Values Across Quantiles

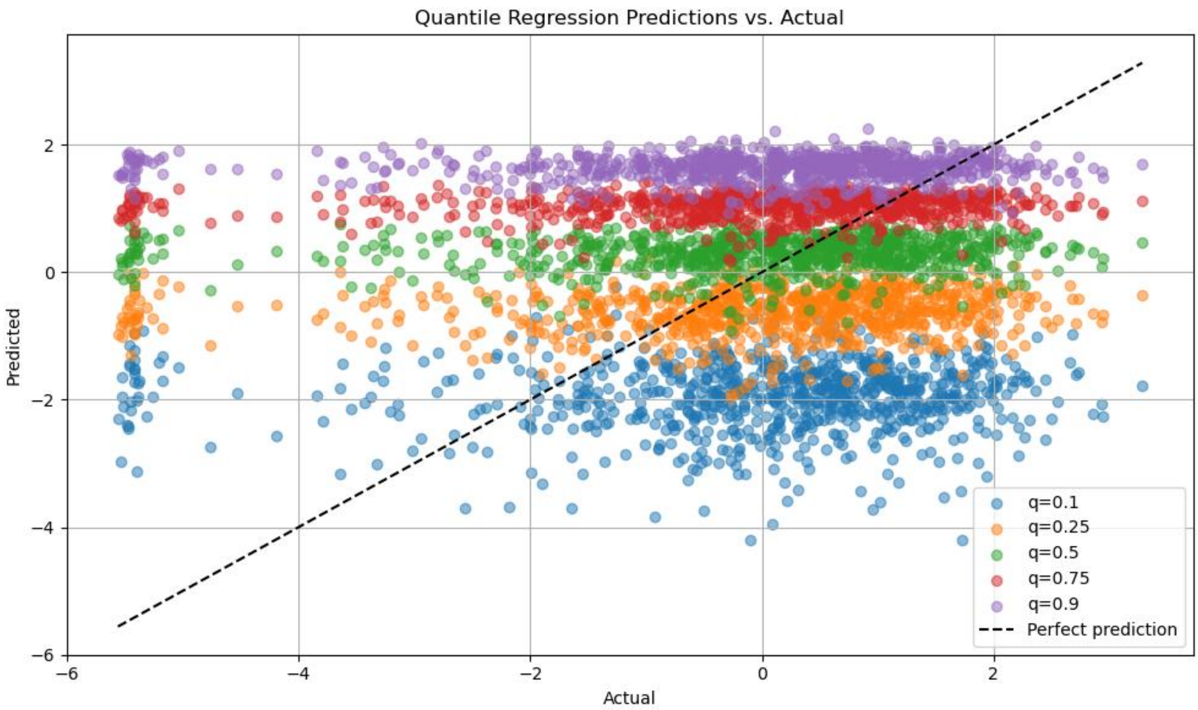


Figure H3. Quantile CatBoost Feature Importances Across Quantiles

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
age_constant_z	12.87	15.40	16.10	14.31	13.09	14.32	15.31	15.82	12.74
number_minors_z	12.87	11.57	10.91	10.94	10.85	11.28	11.35	11.35	10.88
number_adults_z	9.89	11.21	9.86	10.10	10.37	10.03	10.98	10.92	10.64
owns_phone	4.17	2.84	3.44	3.55	3.50	4.25	4.21	3.67	3.39
why_not_purchase_recoded_Lack of Awareness or Understanding	4.35	3.07	4.30	4.28	3.48	4.24	2.80	2.73	2.95
main_info_source_recoded_Media, Community-Based, and Survey Sources	1.69	3.05	4.40	3.95	3.47	3.51	3.51	4.12	4.11
main_info_source_recoded_Professional and Organizational Sources	4.37	3.47	3.23	3.16	2.96	2.81	2.64	2.63	1.67
household_description_Partially settled: Some members of the household are permanently settled and others move around in order to herd their animals	3.44	2.46	3.29	2.98	3.13	2.81	2.71	2.68	2.39
household_moved_Yes	2.14	3.42	2.97	3.26	2.04	2.68	2.59	3.17	2.77
religion_recoded_Traditional/Wakefata	3.85	2.58	2.06	2.47	2.69	2.75	2.59	2.15	3.16
agric_land_Yes	3.45	3.45	2.78	3.20	2.35	1.34	2.36	2.43	2.61
expend	1.76	2.47	3.63	3.08	2.86	2.85	2.34	2.20	2.16
trust_vip_Yes	1.27	2.17	2.22	2.77	2.46	3.67	3.65	2.99	2.15
activity_child_recoded_Working with Livestock	2.32	2.82	1.81	2.77	2.70	3.00	2.01	3.13	2.66
why_not_purchase_recoded_Financial, Practical, and Situational Constraints	3.72	2.73	2.28	1.37	2.06	2.01	2.09	2.05	2.92
why_not_purchase_recoded_Distrust or Negative Perception	1.86	1.64	1.33	1.91	2.95	2.47	2.36	2.94	1.64
religion_recoded_Muslim	1.89	1.26	1.74	2.69	2.29	2.27	2.02	2.07	1.81
activity_child_recoded_Student	2.01	1.56	2.21	1.91	2.57	1.86	1.73	1.53	2.32
educ_recoded_constant_Never attended	2.10	1.65	2.06	2.42	2.08	1.67	1.56	2.29	1.75
educ_recoded_constant_Nursery	2.48	1.48	1.67	1.40	1.69	1.90	2.13	2.57	2.21
know_vip_Yes	1.28	1.62	1.72	1.73	2.47	1.72	1.60	2.14	3.24
activity_child_recoded_Not working	2.28	2.34	1.69	1.65	1.68	1.58	1.83	1.52	2.43
afm_language_Yes	2.14	1.45	1.55	1.53	1.65	1.85	2.65	2.16	1.67
educ_child_recoded_Never attended school	1.99	2.63	2.46	1.86	1.59	1.63	1.60	0.92	1.34
educ_recoded_constant_Elementary	2.84	1.79	1.63	1.54	1.25	1.51	1.37	1.05	1.71
educ_child_recoded_Secondary	1.75	1.19	1.34	1.32	1.85	1.34	1.66	1.39	2.02
eng_language_Yes	0.92	0.85	1.80	1.79	2.05	1.51	1.51	0.97	2.33
educ_recoded_constant_Secondary and Higher	0.93	2.17	1.26	0.89	1.41	1.72	1.90	1.18	1.89
irrigated_land_bin	1.31	1.56	1.65	1.41	1.23	1.02	1.20	0.80	1.30
amh_language_Yes	1.05	1.43	0.75	1.10	1.40	1.53	1.03	1.07	1.17
household_description_Not settled: The whole household moves around in order to herd the animals	0.30	0.72	0.99	0.96	1.01	1.35	1.33	1.54	1.78
educ_child_recoded_Other	0.28	1.32	0.59	1.14	1.89	0.79	0.74	1.25	1.35
activity_child_recoded_Working	0.43	0.61	0.27	0.54	0.94	0.71	0.62	0.56	0.83

Figure H4. Predicted vs. Actual Outcome Values across Quantiles for Quantile CatBoost