AMERICAN UNIVERSITY OF BEIRUT

PREDICTORS OF NEGATIVE COPING MECHANISMS IN UKRAINIAN HOUSEHOLDS DURING CONFLICT: A MACHINE LEARNING APPROACH

by
GHIA NADEEM SANJAR

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Science
to the Graduate Program in Computational Science
of the Faculty of Arts and Sciences
at the American University of Beirut

Beirut, Lebanon
August 2024

AMERICAN UNIVERSITY OF BEIRUT

PREDICTORS OF NEGATIVE COPING MECHANISMS IN UKRAINIAN HOUSEHOLDS DURING CONFLICT: A MACHINE LEARNING APPROACH

by

GHIA NADEEM SANJAR

Approved by:

Dr. Khalil El Asmar, Assistant Professor — Advisor

Department of Epidemiology and Population Health

Dr. Rida Assaf, Assistant Professor — Member of Committee

Department of Computer Science

Dr. Fouad Fouad, Associate Professor of Public Health Practice

Member of Committee

Department of Epidemiology and Population Health

Date of thesis defense: August 27, 2024

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my thesis advisor, Dr. Khalil El Asmar, for his continued support and guidance throughout this research. The path of my research has been greatly influenced by Dr. Khalil El Asmar's perceptive approach of integrating Artificial Intelligence into the healthcare field.

I extend my sincere appreciation to the members of my thesis committee, Dr. Rida Assaf, for his time, invaluable guidance and contributions to the machine learning components of this research. I am truly appreciative of Dr. Fouad Fouad for his thoughtful feedback and deep knowledge of public health, that has helped refine and strengthen the quality of this research.

To my parents, Nadeem and Shafa, I am grateful for your unconditional love, support, and constant encouragement. To my sisters, Layan and Tala, your support, laughter, and advice has given me comfort and encouragement throughout this path. I would have not reached this milestone without you.

Abstract of the Thesis of

Ghia Nadeem Sanjar for Master of Science
Major: Computational Science

Title: Predictors of Negative Coping Mechanisms in Ukrainian Households During Conflict: A Machine Learning Approach

Ukrainian households found themselves forced to adopt negative mechanisms for coping in order to endure the hardships of war in Ukraine. These strategies, including reducing necessary expenses or selling assets, may have long-term effects on household's capacity to effectively handle money and maintain good health. This thesis explores the use of advanced machine learning (ML) algorithms to identify the primary predictors of the harmful coping strategies that Ukrainian households used throughout the conflict. Furthermore, it investigates the effect of several variables, such as proximity to frontline, financial and economic status, and household characteristics, on the likelihood of utilizing negative coping techniques. After finding the optimal set of features for each outcome, Lasso model was used to assess magnitude of the association found in the complex imbalanced dataset between the predictors and the outcomes. Paired with cost-sensitive learning, we were able to identify the significant predictors of negative coping mechanisms. The associations uncovered between the predictors and outcomes revealed two distinct early warning patterns for policymakers: those that serve as protective factors and those that serve as risk factors. These insights, in turn, serve as early warnings for policy makers to target vulnerable households, helping to prevent resorting to negative coping mechanisms.

Keywords: Machine Learning, Negative Coping Mechanisms, Conflict, Ukraine, Household Vulnerability, Imbalanced Data

Table of Contents

Illustrations

Tables

Abbreviations

AAP	Accountability to Affected Populations
ACAPS	The Assessment Capacities Project
ATB	Attention Transformer Block
BN	Batch Norm layer
CV	Cross Validation
FANTA	Food and Nutrition Technical Assistance Project
FAOUN	Food and Agriculture Organization of the United Nations
FC	Fully Connected Layer
FCS	Food Consumption Score
FN	False Negative
FP	False Positive
FTB	Feature Transformer Block
GIS	Geographic Information System
GLU	Gated Linear Unit activation
HH	Household
HHS	Household Hunger Scale
HRW	Human Rights Watch
IDP	Internally Displaced Person
IOM	International Organization for Migration
ISO	International Organization for Standardization
k-NN	K-Nearest Neighbors
LASSO	Least Absolute Shrinkage and Selection Operator
M	Mask
ML	Machine Learning
MSNA	Multi-Sector Needs Assessment
NaN	Not a Number
NGO	Non-Governmental Organization
NFI	Non-Food Item
NLP	Natural Language Processing
NN	Neural Network
OCHA	United Nations Office for the Coordination of Humanitarian Affairs
rCSI	Reduced Coping Strategy Index
ReLU	Rectified Linear Unit activation
RF	Random Forest
ROC AUC	Receiver Operating Characteristic Area Under the Curve

SMOTE	Synthetic Minority Over-Sampling Technique
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
UA	Ukraine
UN	United Nations
UNHCR	United Nations High Commissioner for Refugees
UNICEF	United Nations International Children's Emergency Fund
WASH	Water, Sanitation and Hygiene
WFP	World Food Programme
XGB	XGBoost

Chapter 1

Introduction

1.1 Introduction

In 2022 Ukraine became the center of an intense escalation, since the start of Russo-Ukrainian war back in 2014, bringing with it far-reaching consequences on civilians, particularly households. It has been reported by UN's Office for the Coordination of Humanitarian Affairs (OCHA) that around 14.6 million Ukrainians are in need for humanitarian assistance in 2024 [1]. The increasing severity of military confrontations, border disputes and accompanying humanitarian crisis have been testing the perseverance and resilience of Ukrainian families. Wars heavily affect the living standards of households, leading to dynamic shifts in their welfare status.
It is of critical importance to realize the numerous difficulties that Ukrainian households face due to conflict. Increasing warfare threats force Ukrainian families to flee their homes resulting in a substantial number of internally displaced persons [2]. Often, having to relocate involves losing possessions and sources of income. Economic complications emerge causing job loss, high unemployment rates, and declined purchase power due to change in the exchange rate [3]. Being internally displaced along the expansion of conflict zones has affected children's educational path [4]. Damage of the infrastructure created problems within households by limiting access to necessities like electricity and clean water [5]. Lastly, Doctors without Borders stresses the danger of continuous exposure to violent acts leading to mental health issues such as anxiety and trauma. Also note that access to essential health services might be reduced [6].
The significant impact of the current situation forces families to take haphazard decisions that provide immediate relief. It is imperative, in several domains, to understand the triggers that cause households to resort to such decisions. By identifying these triggers, humanitarian organizations and legislators can design efficient interventions aimed at meeting the needs of households and facilitate proactive preventative measures for upcoming challenges. For instance, understanding the psycho-social roots of these coping strategies can help in targeting societies with mental health interventions to minimize future challenges that might hinder recovery. UNICEF stresses the need for humanitarian involvement to protect children after withdrawing from schools due to closure or damage caused by the war.

Withdrawing children from schools will make them more vulnerable to child labor and violence [7]. These continuous threats affecting the Ukrainian households socially, economically, and mentally call for urgent interference. Despite the fact that this humanitarian issue has been thoroughly studied, we propose an analytical approach which adopts a novel method by making use of advanced machine learning algorithms. These algorithms will improve our comprehension of the association between the harmful coping mechanisms and the factors in our study such as proximity to conflict and several household characteristics. This tool will shed light on unnoticed trends and influential factors that might have been overlooked in the conventional humanitarian research. The integration of this innovative approach in this research represents a significant opportunity to create an effective framework to address the needs of affected households and lessen the long-term effects on the targeted households through focused interventions. The objectives of this research include:

- To determine the factors that drive unhealthy coping strategies in Ukrainian households during hostilities.
- To explore the effects of several factors on the adoption of negative coping methods.
- To contribute to the understanding/knowledge of long-term impact of negative coping mechanisms on war-affected households.
- To uncover patterns and associations related to negative coping mechanisms using advanced machine learning techniques.

1.2 Literature Review

By coping mechanisms, we mean the process of managing responses to situations of conflict and distress through actions and thoughts [8]. Folkman and Moskowitz (2004) named four categories of coping mechanisms across various literature:

1. Problem-focused
2. Emotion-focused
3. Meaning-focused
4. Social coping

While all strategies are important, problem-focused and social coping are the most relevant to this research. Problem-based coping is minimization of distress through one's behavior [9] and social coping is minimization through community support like supplying tangible needs [10]. Certain strategies are considered negative because, although they provide instant comfort and relief in times of distress, they have detrimental long-term effects on household's wellbeing. Negative coping mechanisms relevant to this research and encountered in the literature are the following:

1. move elsewhere [11]
2. sell household assets [12]
3. borrow food [13];[14]
4. eat elsewhere [13]
5. sell productive assets [15]; [16]; [12]
6. reduce education expenditures [17]
7. sell house/land [12]
8. asking strangers (social support) [12]
9. reduce health expenditures [18]; [17]
10. spend savings [17]
11. degrading income source [12]

Consider a family faces severe financial stress due to conflict in the area. One potential response would be to reduce education expenditures, like tuition fees and other associated costs, by withdrawing children from school. Bringing impermanent financial relief to the family, they overlook the collective impact on the child's future economic and social potential. Without proper education, they will encounter limited employment options [19], reduced future wages [20], feel alienated accompanied with loss of self-esteem, lack social skills and limit innovation and problem-solving skills. Consequently, an uneducated community will not have the tools to conquer socio-economic challenges.
Various research agrees that the degree of employing negative coping strategies relies on economic repercussions of the conflict, severity of conflict, and household's economic and demographic characteristics. Some of these characteristics consist of house welfare, level of human capital, gender, financial status, asset ownership, employment status, income sources, education of household head, internal displacement, and proximity to conflict [21]; [22]; [23]; [24]; [25]; [26]. Moreover, Palma and Araos (2021) stress that the ability to cope properly lies within the potential of identifying outside limitations and allocating resources appropriately. All households, whether in combat or non-combat areas, are affected by the war since institutional changes occur on economic and social levels [23]. Distance from frontline is one of many predictors for assessing severity of utilizing coping mechanisms such as moving elsewhere. household members are either internally displaced i.e., move places within the country to a safer environment, or move to another country to ensure better living conditions and secure an income [27] [28]. In Ukraine, around 5.1 million are internally displaced (IDPs) as of May 23rd, 2023, originating from Kharkiv, Donetsk, Zaporizhya, Kherson and Luhansk [2]. Located in Eastern Ukraine, these oblasts are considered danger zones since some are under the Russian federation

military control and contain landmine sites [29]. According to REACH, IDPs arriving at transit sites, like Dnipro, are children, women, elderly, and disabled persons. Often, these come from a poor socio-economic background [30].
Households, failing to accumulate assets, secure an income or borrow, resort to spending their savings [23]. Justino (2011) emphasizes that the vulnerability of household welfare is dependent on their savings. In accordance, Hill et al. (2019) point out that the poorer the household the more likely to resort to such coping strategy. A more detailed study states that the wealthier the household the more likely to use savings and sell assets but are less likely to borrow [26]. Tongruksawattana et al. (2010) continues to say that the higher the income of a household accompanied with asset loss from economic shocks, the more likely they resort to borrowing and not spending their savings. On the contrary, research on civil war in Southern Sudan pointed out that the vulnerability of households during these challenging times does not rely on their financial status [21]. Some households hold onto productive assets, such as cattle, livestock, and farmland, to sell in times of conflict. However, Justino (2011) argues that livestock are prone to thefts or death in risky environments. An economic literature, on household coping in Rwanda, compared the cattle sales during wartime and peacetime and found that sales decreased during wartime [16]. More specifically, Verpoorten (2008) argues that selling productive assets is ineffective in areas witnessing violence. An interesting gender analysis by IOM on internal displacement in Ukraine reported that females utilized savings to secure basic household needs while males mainly relied on selling household goods. Moreover, they state that the employment proportion among IDP (40%) is less than the non-displaced population [2].
Expenditure minimizing coping strategies include reducing health and education expenditures. Two studies on the association between armed conflicts and utilization of health care services showed that financially struggling households that live in affected areas limit their access to such services [18]; [31]. In a health needs assessment report, REACH reported health related challenges of households in Dnipro and Zaporizka oblasts and Kharkiv city. Households report financial challenges when accessing medical care and medications. More specifically, households in Kharkiv city reduce other expenditures in order to afford medication [32].
As for education, households choose to minimize education expenditures to lessen the financial strain experienced [33]. Arkesh and De Walque (2008) were able to demonstrate that the detrimental effects on education systems are more pronounced in the provinces with higher rates of genocide [20]. So, the closer to areas of conflict the more likely the education process is disrupted due to safety issues and damage to facilities. However, UNICEF are making noticeable efforts to give Ukrainian children access to education. Since the beginning of the war UNICEF assisted approximately 1.5 million children to access education, distributed development kits and supported Ukrainian school through $3 million in cash [7].
Coping Strategies Index, an indicator of household food security, lists consumption coping mechanisms adopted by households. The first is borrowing food which increases food availability temporarily; the second is sending household members to eat elsewhere; the last is changing their food intake (diet) to cheaper less preferred

items [13]. According to a survey by WFP, 70% of interviewed Ukrainian households adopted common food coping strategies such as eating less preferred or less expensive food. Testimonies from Ukrainians residing in Zakarpattia indicated that they no longer buy in big quantities, and they rely on food from humanitarian organizations to save money [34].
ACAPS Yemen Analysis Hub researched Yemeni households' experiences and explored coping strategies they adopted during war. They found that lack of basic services such as water, electricity, and nutrition related materials, drives communities to cooperate on filling the service gaps. Households extend support through sharing water tanks, gas for cooking, stored energy during power cuts, and food. While receiving cash assistance from NGOs may seem to aid, ACAPS highlights that it could be of little benefit to the families considering debt repayments and basic commitments such as food and sending children to school. The analysis accentuates that Yemenis tend to sell their assets such as land and gold as a last resort when income is no longer available or enough to cover the cost of living [12].

1.3 A Machine Learning Approach

Research in active war sites is often hard to manage due to several factors such as communication barriers and violence. Besides, data continually varies according to the country's situation and following-up households on a temporary basis is quite impossible. Machine learning can offer insights and adaptations that traditional approaches might overlook [35]. The use of traditional statistical methods requires some assumptions, such as type of error distribution, to be met before proceeding to avoid prediction errors and misinterpretation of results [36]; [37]. Using these methods require an in-depth knowledge about the robustness of models to assumption violation. Researchers should be aware of the underlying model that produced the data. However, real-life data is inherently complex, dynamic, and non-linear. Therefore, to extract useful insights, machine learning (ML) algorithms designed to adapt to data with no prior assumption of a model are used [38]. Instead of depending on pre-existing algorithms, it uses the data to construct and continuously improve a predictive model [39]. The importance of these models lies in their scalability, which is the capability of handling large amounts of data and complex interactions between features [38]. ML algorithms excel at revealing non-linear relationships and complex patterns that simple models might overlook in large datasets. On the other hand, classical models are only useful when data is not large, and the relation between the variables is stable i.e., assume stationary distributions [39]. Thus, comparing machine learning to conventional statistical models reveals distinct advantages that set it apart as an effective and flexible method in various domains.
The data used in this research is imbalanced where households actually using negative coping mechanisms are small in proportion. Conventional statistical models perform poorly on imbalanced data since they assume the equal representations of both classes in binary classification [40]. This leads to a misleading high accuracy measure. ML techniques are flexible in offering adaptations, such as class weighing and parameter tuning that can handle evolving datasets [37]. For instance, given

the warfare setting, our data is unstable where it heavily depends on the accessibility of households and the level of intensity of war. However, ML models can generalize well on unseen data through the previously identified characteristics if tuned and evaluated correctly. In conclusion, applying advanced machine learning techniques offers a viable means of outperforming conventional statistical models in terms of more precise analysis, deeper understanding of underlying patterns, and flexible modeling capabilities.

Our study aims to explore the predictors of negative coping mechanisms employed by households during warfare using ML techniques and compare our results with those found in the literature. In addition, prevalence predictions of coping strategies will help in assessing household resilience. This sheds light on the need for early intervention and humanitarian assistance to mitigate severity and equip households with skills to overcome future challenges. Several papers present similar approaches that we will adopt later in the study. The first identified the predictors of psychological distress caused by COVID-19 by using a random forest model, which outputs the importance of predictors [41]. The second compares multiple machine learning models, which predict household resilience towards stressors and shocks using the best set of selected predictors [42]. Another study exploring the impact of armed conflict and natural disaster on societal resilience toward climate hazards, suggested the use of machine learning techniques over conventional models. To determine the features that provide the best explanation for the target variable and establish associations, random forest analysis was used [43]. D'Angeli and Vesco (2022) argue that armed conflicts are indeed associated to societal vulnerability towards future hazards. However, this study does not delve deep into the underlying stressors caused by armed conflicts and natural disasters in society. Furthermore, Garbero and Letta [42], used multiple ML classification models such as bagging, k-NN, SVM, and random forests to predict household resilience. They listed features that significantly contributed to household resilience. This is similar to our approach of identifying stressors of negative coping mechanisms during wartime in Ukraine. However, their setting was focused on households experiencing shocks and not explicit to households during wartime. To our best of knowledge and the literature found, there are no studies empirically investigating the impact of war on households' resilience in terms of employing negative coping mechanisms through advanced analytics.

Chapter 2

Methods

2.1 Data Description

World Food Programme (WFP) and REACH Initiative worked together to conduct a cross-sectional household-level survey know as Ukraine Multi-Sector Needs Assessment (MSNA). WFP is an international organization within the United Nations that provides food aid worldwide. REACH is a leading humanitarian initiative offering precise data, up-to-date information, and thorough analysis from crisis, disaster, and displacement contexts [44]. Its work directly contributes to aid response and decision-making by offering precise and easily accessible information on the humanitarian conditions of populations affected by crises [45].

A representative sample of $33,283$ household members were surveyed through face-to-face and phone interviews between October and December 2022. These individuals represent a total of $13,322$ Ukrainian households. Interviews were held in 55 raions across 22 oblasts. Raions are administrative subdivisions within oblasts, which are larger regional divisions similar to states or provinces.

The data reflects the situation in accessible areas at the time of the survey only. Findings on the inaccessible areas such as Kherson and Luhansk oblasts are found in a separate assessment report [46].

Our data consists of three separate datasets: one contains household level data, the second contains household member characteristics and the third is health care related at the household member level.

Among these datasets, we come across 12 themes. They are as follows:

- **Demographics** include household characteristics such as gender, employment, etc..., and Washington Group questions that help identify persons with disabilities.
- **Displacement** tracks household member separation, location and movement.
- **Education** tackles education accessibility, learning modalities and barriers.
- **Shelter & NFI** tackles the living conditions of households.
- **Emergency Telecommunication** tackles internet access and coverage.

- **WASH** tackles hygiene and water accessibility, sanitation and sufficiency.
- **Food Security** includes well-known scales that assess food security such as Food Consumption Score, Household Hunger Scale and Reduced Coping Strategies Index.
- **Cash & Markets** covers incomes, expenditures and monetary challenges.
- **Livelihoods** which are the target variables in this research. It informs us of the use of livelihood coping strategies.
- **Health** tackles healthcare services, medicine sought by household members and obstacles faced in terms of accessibility.
- **Protection** includes security, property, children and women safety concerns, and barriers in accessing governmental and community services.
- **AAP** tackles assistance related information and records household satisfaction.

2.2 Member-level Data Transformation

The member-level data was transformed into household-level data for us to use in the analysis. Our aim was to capture as much as possible information about the members in each household. We chose to use data aggregation by the unique household id for the column transformation process. Each feature in the member-level data was transformed into a corresponding household-level feature by translating it into the sum (count) of every category in a feature. For instance, take the gender feature which labels each member as either male or female. At the household level, this feature was split into two distinct columns, one for the number of females and one for the number of males in the household. This applies to all the features in the data. This approach allowed us to aggregate individual member characteristics into meaningful household-level metrics, thereby providing a complete view of household dynamics. After we transform all the columns, the household-level data and the transformed data were merged by household id resulting in a dataset of almost 800 features.

2.3 Data Preprocessing

A systematic recoding process was employed since the data consists of mainly categorical variables. Features with response indicating negation or options such as "no", "prefer not to answer", "don't know" were coded as $'0'$, otherwise they were coded as $'1'$. Missing values in the dataset arise from skip patterns inherent in the survey design and therefore they are not random. This means that certain questions were only directed at specific groups, for instance a particular age group, resulting in skipped responses for those who do not meet the criteria. These skipped responses

were coded as ′98′ to differentiate them from other types of missing data. Later when modeling, these skipped values were recoded to ′0′ since the individuals were not eligible to respond, indicating no positive outcome.

Furthermore, specific variables such as the Household Hunger Scale (HHS), reduced Coping Strategy Index (rCSI), and Food Consumption Score (FCS) were recoded using procedures provided by the World Food Programme (WFP) and the FANTA Project. For continuous variables, indicators of distance from the frontline, discussed in appendix A, were incorporated into the dataset, and income and expense indicators were computed by aggregating there values.

Once recoding was completed, the missing values were imputed. For categorical variables, missing values were imputed using the mode, while for continuous variables, the mean was used. These imputations were conducted within each raion (region) to ensure that the imputations are consistent with the characteristics of households in the same region. Categorical features were one-hot encoded to reduce collinearity in model and to ensure interpretability of the Lasso model later in the research. Numerical features were standardized to improve model performance, interpretability and consistency across features.

Lastly, an approach we took to reduce model complexity furthermore was dropping the columns where over 90% of its responses are '0'. These column offer little variability, can affect the model's power negatively and require a high computing efficiency. By reducing the number of features, we simplify our model, use meaningful features and potentially reduce the risk of overfitting.

2.4 Class Imbalance

Class imbalance is a common problem in machine learning specifically in real-world applications. It occurs when the number of instances in one class is significantly lower compared to the number of instances in another class. In other words, it's the case where the outcome is a rare event. The negative coping mechanism outcomes show significant class imbalance where the percentage of households who employ negative coping strategies is much lower than those who do not. Figure 2.1 shows the proportion of minority class for each negative coping mechanism available in the data. The majority of the target variables are moderately imbalanced with the percentage of the minority class ranging from about 1.28% to 15.98%. However two features, *sell house* and *sell productive assets* with minority class proportions of just 0.18% and 0.68%, respectively, stand out for their extreme imbalance [47]. This large difference suggests that these two behaviors are not common nowadays. Since the minority class instances are severely underrepresented, such high imbalances may affect the models' performance and requires the use of specialized strategies to successfully handle the imbalance and eliminate any bias.

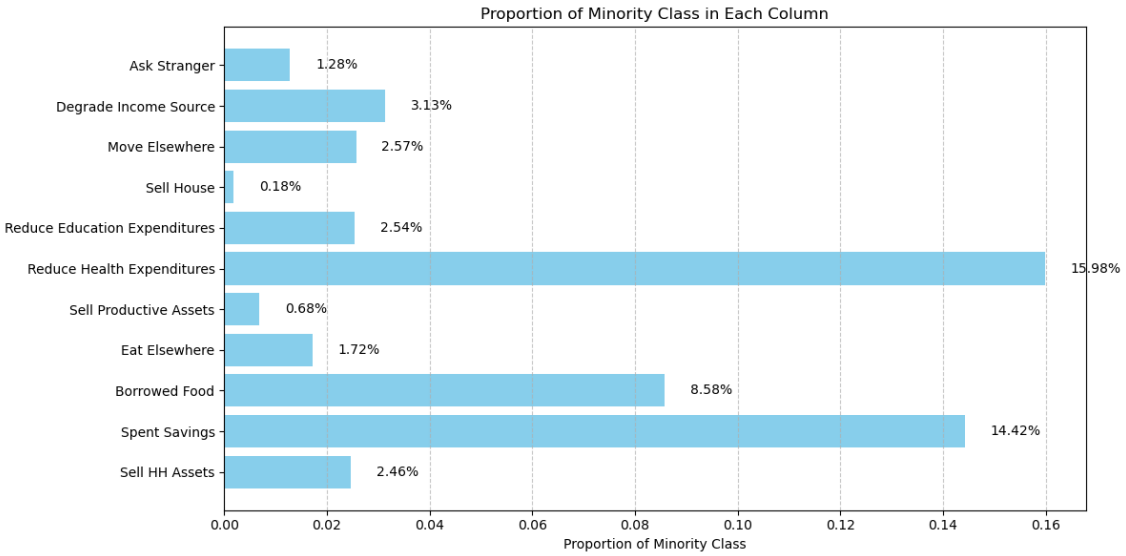


Figure 2.1: Proportion of Minority Class for Each Negative Coping Mechanism

2.5 Approach to Handling Imbalanced Data

Predictive classification models, such as decision trees or logistic regression, are often biased towards rare events. These models often struggle with misclassifying the minority classes as they consider them noisy data and less representative. The model's goal is to accurately predict by reducing the loss i.e., minimize false positive and false negative predictions. Thus, if not accounted for class distribution, the predictive models are unreliable for reporting inaccurate and biased results. There are several approaches, which can be used separately or together, to deal with class imbalance. The first approach consists of sampling techniques like oversampling or undersampling. The second approach is using synthetic data generation where synthetic samples of the minority class are created using SMOTE (Synthetic Minority Over-Sampling Technique). Thirdly, use a weighted class approach that involves assigning higher weights to the minority class observations so the model will prioritize correctly classifying them. This approach is called the cost-sensitive learning. For the first part of our research, where we intend to do features selection, we employed undersampling, also known as downsampling, to create a smaller balanced data without the use of synthetic data. Undersampling is the process of reducing instances of the majority class $N_{DidNotEmploy}$. Randomly sample $n_{DidNotEmploy}$ such that $n_{DidNotEmploy} = N_{employed}$ of those who employed the negative coping mechanism [48]. In addition, cost-sensitive learning was utilized in the random forest model where the class weights were adjusted. For the second part of the research of quantifying the associations between the target variables and features using Lasso, we use a cost-sensitive approach on the imbalanced dataset where the class weights are also adjusted.

2.6 Distance Features

Frontline Map Representation [49]
Throughout the literature, proximity of households to the frontline was found to have a direct effect on the household's welfare. To be able to assess this relationship, we make use of geo-spatial data which are designed to work on a location component. One particular type of geo-spatial data that defines the precise location and size of a country's territory is geo-spatial data with country boundaries. The administrative boundaries, in this research, are *Oblasts* and *Raions*. In short, we calculated the distance from the household's location to the raions located on the frontline. Check Appendix A for details on GIS data used in this research.

2.7 Modeling design

One of our goals was to detect features that indicate the use of negative coping mechanisms using advanced analytics. For each target variable, several models mentioned in the following section were trained, tuned, and eventually evaluated using stratified k-fold cross validation. Furthermore, resampling techniques, such as undersampling, as well as class weight-balancing methods were employed when using the models. We used eight negative coping mechanisms from the Ukraine MSNA data and for each we provided a focused analysis offering intricate insights that have not been explored before in the context of conflict. We fit four predictive models for each of the outcomes. Their corresponding model performance metrics, presented in the next chapter, display the baseline (untuned) and tuned model metrics for the sake of showing that tuning is an important step for a better predictive results. Also, the $ROC - AUC$ curves of the well-performing models were displayed. This process was repeated for each negative coping mechanism separately.

2.8 Predictive models

2.8.1 *Random forest:*

Random forest (RF) is a meta estimator that employs averaging to increase prediction accuracy and manage over-fitting after fitting several decision tree classifiers on different subsamples of the dataset [50]. It is important to note that RFs are bagging techniques, which is a remedy to class imbalance. There are more ways that RFs address the issue of class imbalance. In this research, we adjust the class weights to balanced sub-sampling accompanied with undersampling.

2.8.2 *Gradient Boosting Algorithms:*

XGBoost is a popular and widely utilized boosting ensemble technique renowned for its efficiency and adaptability. It is favored due to its scalability compared to other boosting algorithms, handling imbalanced data, and providing feature importance.

2.8.3 *Lasso-logistic regression:*

Introduced by Tibshirani in 1996 [51], Lasso is an extension of logistic regression approach by applying the L_1 regularization on the coefficient vector β. The Lasso penalty function improves the model's capacity to identify patterns in the minority class [52]. It produces a sparse coefficient vector β with the zero entries being non-contributing features.
The Lasso penalty function:

$$P(\beta) = \sum_{j=1}^{p} |\beta_j| \quad \text{with} \quad \sum_{j=1}^{p} |\beta_j| < \lambda$$

The log-likelihood of Penalized logistic regression using Lasso penalty:

$$log(L_p(\mathbf{y}; \beta)) = \sum_{i=1}^{n} -[y_i log(\pi_i) + (1 - y_i) log(1 - \pi_i)] + \lambda \sum_{j=1}^{p} |\beta_j|$$

2.8.4 *TabNet:*

Neural Networks (NNs) are the go to methods for many topics like NLPs and computer visions, however conventional approaches, such as boosting techniques, outperform them when handling tabular data. As a result and to fill the gap between boosting algorithms and NNs, without loosing the explainability factor, the TabNet model was created. TabNet, which is a deep learning network, was created specifically to handle tabular data [53]. It can deal with structured data with different column types and data dimensions since it combines the advantages of deep learning techniques with tabular data representation. In order to find and concentrate on the most important features during training, TabNet makes use of a unique attention-based sparse feature selection approach. TabNet is an effective tool for analyzing tabular data because of its feature selection process and adaptive sequential decision-making technique, which enables it to perform effectively on tasks like time series forecasting, regression, and classification [53].

How does TabNet work?

In a brief overview, it consists of an attention transformer block, feature transformer block, and sequential steps. The model takes in raw features as an input which enter through a batch norm (BN) layer followed by a feature transformer block (FTB). In TabNet, FTB has a gated linear unit activation 'GLU' which helps in sharing layers within the architecture of the model. Then, the attention transformer block (ATB) creates a Mask (M) which allows instance-wise feature selection for explainability and efficient learning capacity. After that, the features are masked for the next block of feature transformer. At each step, we note two things happening inside the global TabNet architecture (refer to figure 2.2). This first is that there's a prediction happening at each step where the decision passes through the 'ReLU' and the predictions are being summed up. The second is that at each step the Mask is used

to get the feature attributes which are then aggregated. In general, having these sequential steps resembles the ensemble methods to enhance the model's performance. These steps are repeated n times until finally reaching the fully connected layer (FC) suitable for our binary classification problem as well as extracting the global feature attributes. For better understanding of the FTB and ATB mechanism, refer to figures 2.3 and 2.4.

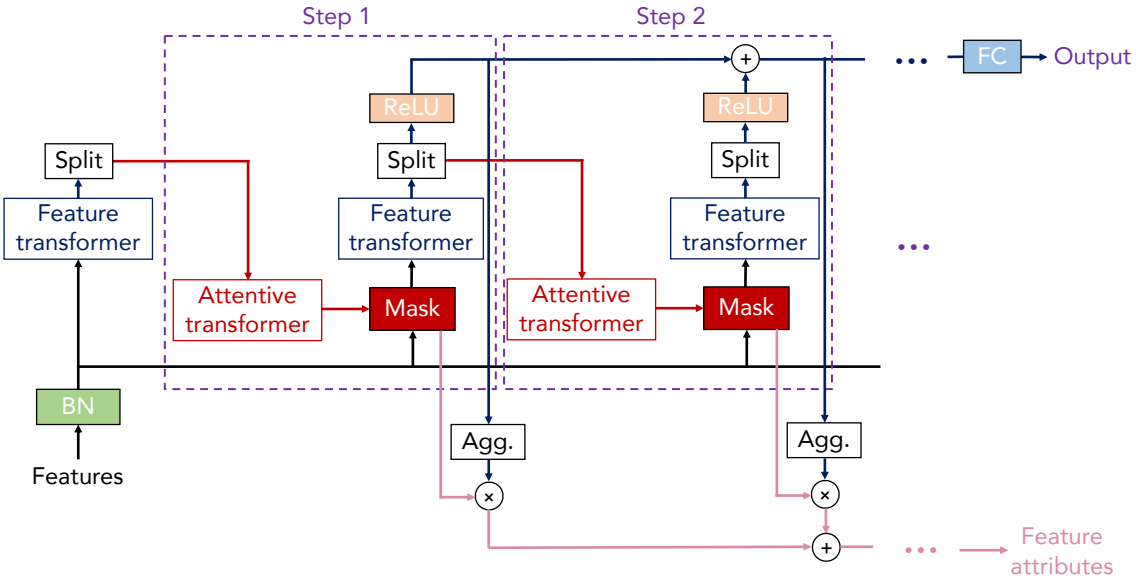


Figure 2.2: TabNet Global Architecture [53]

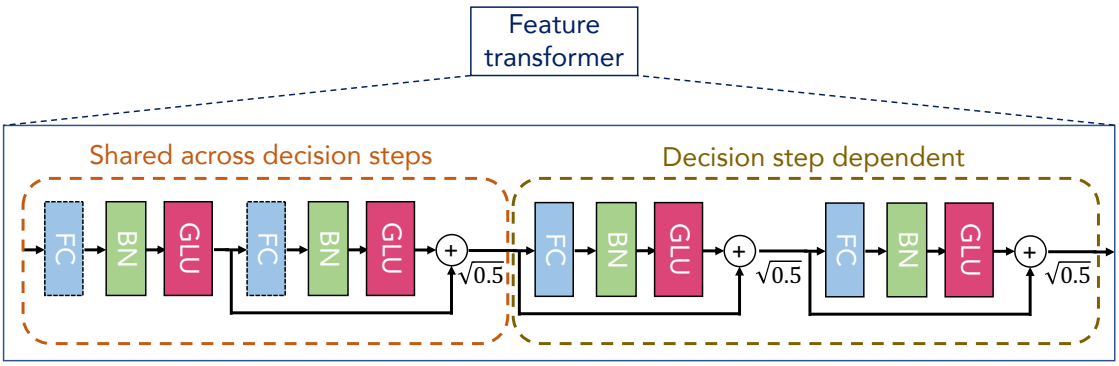


Figure 2.3: Feature Transformer Mechanism. It shows 4 consecutive blocks each consists of FC $\rightarrow$ BN $\rightarrow$ GLU. The 'GLU' is $\sigma(x) \bigodot x$ (sigmoid function $\times$ input features). At each step, the first two blocks are going to be shared meaning that the model is sharing weights in between steps. The skip connection at each block, allows a deeper and smoother learning procedure [53].

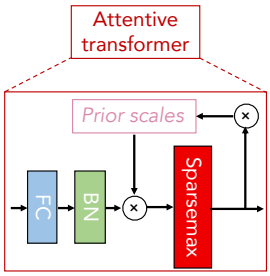


Figure 2.4: Attentive Transformer Mechanism. It consists of a FC $\rightarrow$ BN followed by prior scales P which are initialized as $P_0 = 1$. At each step i, the prior steps are computed as follows $P_i = \prod_{j=1}^{i}(\gamma - M_j)$. Prior scales give out information on the features and how much of them is used. When γ is close to 1, the model used different features at every step. When γ is large, features can be repeated at every step. The *sparsemax* [54] function allows instance-wise feature selection. It outputs probabilities that sum to one and reduces dimensionality [53].

2.9 Feature Selection

Multicollinearity and high dimensional input variables might reduce the model's performance and increase its computational cost. The use of feature selection techniques can narrow down the number of inputs to just those that we think the model will benefit from while maintaining interpretable results. To achieve this, we used embedded methods that apply feature selection within the training process. Random forests, XGBoost and TabNet have a built-in function that ranks the features according to their importance. It is crucial to point out that random forests use bagging ensemble method while XGBoost uses boosting ensemble method. Therefore, both algorithms generated different sets of features that are considered important for explaining the outcome. In addition to this, the Lasso algorithm identifies the optimal set of features associated with the target variable and shrinks the coefficients of unnecessary features to zero. We aim to assess whether these ML models identify the same stressors found in the literature and explore any potential stressors that the literature might have overlooked. The optimal set of features identified for each outcome will be the results of features recognized as important by the models only. The next step would be comparing to those identified by the literature. After extracting the feature importance and ranking them in descending order, the top 10 features of each model were chosen with a threshold of 1%. If a feature has an importance less than 1%, it was dropped. Next, the features were ranked in a bar graph representation according to how frequently they appear in the top 10 features. This means that if a feature occurred within the top 10 in 4 of the models presented, then the bar graph would indicate that this feature occurred 4 times.

2.10 Cross Validation

Stratified 5-fold and 10-fold cross validation were used to assess the models. An advantage of this method over the normal train-test split is that it iteratively trains and tests on k folds maintaining the actual class distribution. In an imbalanced setting, it offers more robust and reliable performance and better assesses the generalization ability of a model [55].

2.11 Evaluation Metrics

Selecting appropriate performance evaluation metrics is essential to ensure successful classification of households using negative coping strategy. Accuracy, often used in classification problems, gives misleading high scores and neglects minority class performance in an imbalanced setting. However, balancing the data using undersampling, accuracy becomes a crucial metric that we aim to maximize. Therefore, the following metrics were chosen to evaluate our models:

1. **Accuracy** It measures the proportion of correctly predicted instances out of the total instances.
   Accuracy is calculated as follows:
   $$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
   Where:
   - TP represents the number of true positives.
   - TN represents the number of true negatives.
   - FP represents the number of false positives.
   - FN represents the number of false negatives.

2. **Precision** It calculates the percentage of positively identified instances that are indeed true.
   Precision is calculated as follows:
   $$precision = \frac{TP}{TP + FP}$$

3. **Recall** Also known as sensitivity, measures the percentage of actual positives correctly predicted by the predictive model.
   Recall is calculated as follows:
   $$recall = \frac{TP}{TP + FN}$$

4. **F-beta Score** Is a metric that controls precision-recall trade-off. It is a robust measure when evaluating model performance in case of class imbalance. It assesses the classifier based on false negatives and false positives.
   $$F_\beta = (1 + \beta^2)\frac{\text{precision} + \text{recall}}{(\beta^2 \times \text{precision}) + \text{recall}}$$

5. **AUC ROC** For binary classification problems, an evaluation metric known as the Receiver Operator Characteristic (ROC) curve is used. It plots the true positive rate against the false positive rate at different threshold values. Stated differently, it displays a classification model's performance across all classification thresholds.

6. **Learning Curve** In machine learning, a learning curve is a graph representation that demonstrates how well a model performs as the training set size increases [56]. In this reserach, the scoring method of learning curve is based on the $ROC\text{-}AUC$ score given the imbalanced setting in this research. Learning curves are useful tools for detecting overfitting or underfitting and determining whether more data, better features, or different methods could improve performance.

2.12 Quantifying the Associations

Our second objective, after identifying the precursors of negative coping mechanisms, was to quantify their relation with respect to the outcomes. Lasso served as a mean to quantify these associations in addition to excluding any unnecessary feature in the model. Features with 0 as their coefficient or have an insignificant p-value greater than or equal to 0.05 were dropped from the model. The remaining coefficients were interpreted as the change in odds of the outcome for one unit increase in the feature. The model was assessed using performance metrics such as $precision$, $recall$, F_1 score and AUC. Moreover, the learning curve with an $ROC - AUC$ scoring method was used to further assess the model. On a final note, the Lasso model was fit on the whole imbalanced dataset with only cost-sensitive learning applied. The class weights of the data were set to *'balanced'* and the cost parameter C was optimized.

Chapter 3

Results

3.0.1 *Target Variables Chosen for the Study*

The selection criteria for these target variables were based on their minority class proportion, which ranges from 1.28% to 16%. The outcome that had a minority proportion of 1% or less, such as *selling household* and *selling productive assets*, were excluded because they are less prevalent in current times. Instead, household tend to employ other strategies before resorting to such extreme measures. The variables we retained for the study include *reduce health expenditures*, *spend savings*, *borrow food*, *degrade income source*, *reduce education expenditures*, *sell household assets*, *eat elsewhere*, and *ask strangers*. These selected variables reflect more relevant coping mechanisms and provide a meaningful basis for analysis.

3.1 Model Selection

In this project, the models that contributed to the feature selection for each outcome were Lasso, Random Forests, XGBoost and TabNet. Each was used with 5-fold and 10-fold stratified cross validation. Consequently, there are 60 different model assessments used in this process. In this research we refer to the 5- fold and 10-fold models as one model since the results of the metrics using cross validation were averaged. Metrics include the *accuracy*, *AUC*, *Precision*, *Recall*, and feature importance. The 10-fold models demonstrated better performance compared to the 5-fold models. Therefore, the result reporting included the 10-fold models exclusively.

3.2 Model Performance for Reduction of Health Expenditures Coping Mechanism using Downsampling

Dataset Dimension: $(4258, 179)$

Table 3.2 refers to the results of *reduce health expenditures* outcome. In the baseline comparison, RF outperforms, achieving the highest $f_1Score = 0.80$, $recall = 0.84$, $accuracy = 79.61\%$, and $AUC = 0.87$. This indicates its robustness in identifying true positives while maintaining a strong accuracy. XGB also performs well, with

Table 3.1: 5-fold stratified cross validated performance using different performance measures and confusion matrix values for Reducing Health Expenditures

Type	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.75	0.74	0.76	0.7529	0.83	1620	543	1586	509
	XGB	0.79	0.81	0.77	0.7858	0.86	**1628**	411	1718	**501**
	RF	**0.80**	**0.83**	**0.78**	**0.7963**	**0.87**	1620	**358**	**1771**	509
	TabNet	0.78	**0.83**	0.73	0.7616	0.82	1477	363	1766	652
Tuned	Lasso	0.75	0.75	0.76	0.7536	0.83	**1618**	538	1591	**511**
	XGB	0.80	0.82	0.77	0.7905	**0.87**	1615	378	1751	514
	RF	**0.80**	**0.84**	**0.77**	**0.7935**	**0.87**	1593	**343**	**1786**	536
	TabNet	0.78	0.83	0.73	0.7604	0.82	1468	359	1770	661

Table 3.2: 10-fold stratified cross validated performance for Reducing Health Expenditures

Method	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.75	0.75	0.76	0.7534	0.83	1611	532	1597	518
	XGB	0.79	0.81	0.78	0.7895	0.86	**1637**	404	1725	**492**
	RF	**0.8**	**0.84**	0.77	**0.7961**	**0.87**	1603	**342**	**1787**	526
	TabNet	0.77	0.80	0.74	0.7595	0.82	1534	429	1700	595
Tuned	Lasso	0.76	0.76	0.76	0.7562	0.84	1610	519	1610	519
	XGB	0.80	0.82	0.77	0.7926	0.87	**1619**	373	1756	**510**
	RF	**0.8**	**0.84**	**0.77**	**0.7940**	**0.87**	1590	**338**	**1791**	539
	TabNet	0.78	0.82	0.75	0.7719	0.82	1539	381	1748	590

high $recall = 0.81$ and $AUC = 0.86$, while its $precision = 0.78$ and $f_1Score = 0.79$ are slightly lower than RF. Lasso and TabNet show weaker performance, with Lasso achieving the lowest F_1 Score of 0.75, an AUC of 0.83 and and $accuracy = 75.34\%$, while TabNet has the lowest score in $precision = 0.74$.

After tuning, RF continues to lead, with consistent f_1Score, $recall$, $accuracy$, and AUC, showing that tuning does not significantly change the model's performance. XGB shows slight improvement in $f_1Score = 0.80$ an its high $AUC = 0.87$, making it a competitive alternative to RF. However, RF outperforms XGB with a higher $recall$ which means that it correctly classifies the negative classes more. Lasso and TabNet improve slightly in f_1Score and $recall$. However, TabNet still lags behind in $precision = 0.75$ compared to the other models.

3.2.1 *Set of Best Parameters*

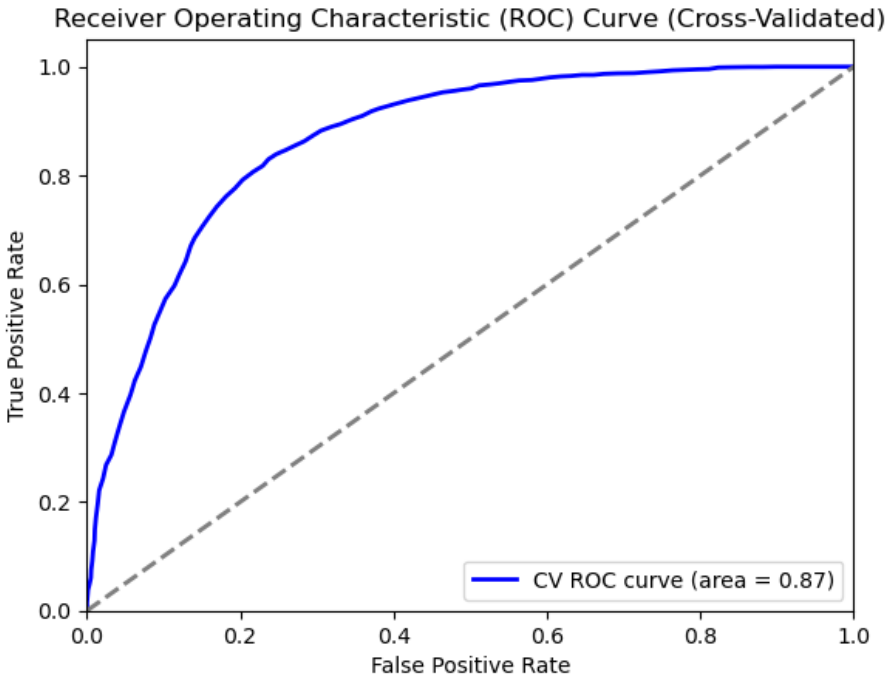


Figure 3.1: AUC-ROC curve of the Baseline Random Forest model with a 10 Fold Stratified CV

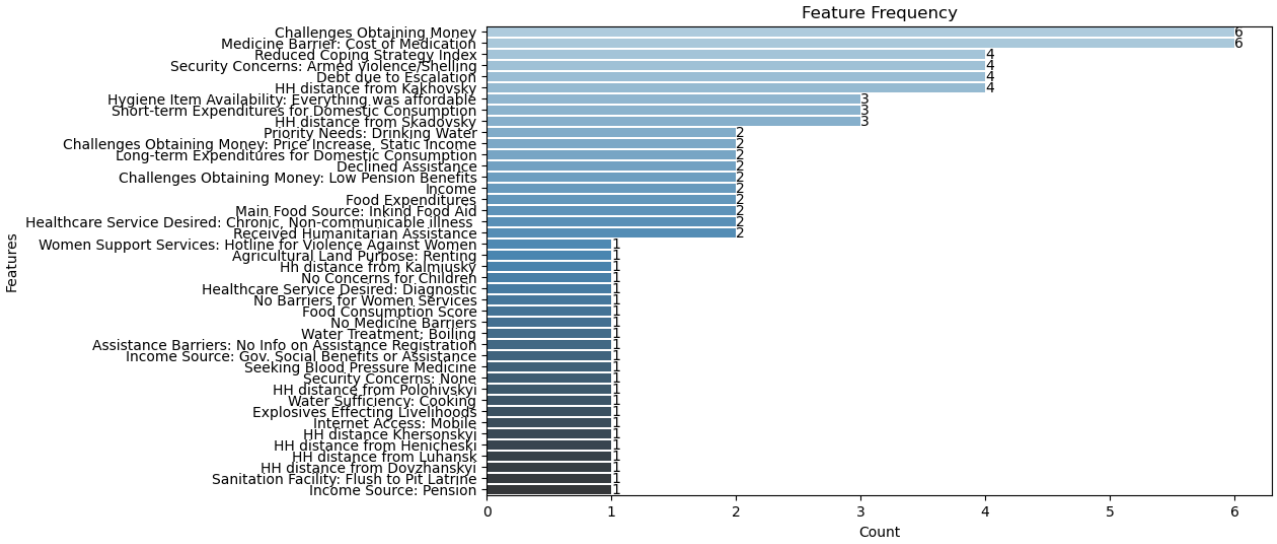


Figure 3.2: Optimal Set of Features: Reduce Health Expenditures

Starting with the *Reduce Health Expenditures* outcome, figure 3.2 illustrates the important features that are possibly associated with the outcome. Features such as challenges obtaining money and cost-related medication barrier appeared in six of the eight models. This reflects consistency between the models on selecting these features as one of the top 10 important features. Two of the models propose a more accurate explanation on the challenge obtaining money feature. Households

Model	Parameter	5-Fold CV	10-Fold CV
Lasso	C	0.1	0.1
	Solver	'Saga'	'Liblinear'
XGB	N_estimators	200	200
	Max Depth	7	7
	Learning rate	0.1	0.1
	Subsample	0.9	1
	Gamma	0.01	0.1
	Alpha	10	10
Random Forest	N_estimators	100	200
	Max Depth	'None'	30
	Min Sample Split	5	2
	Min Sample Leaf	2	1
	Max features	'sqrt'	'sqrt'
TabNet	n_d	48	48
	n_a	48	48
	n_steps	3	3
	Momentum	0.05	0.05
	Clip	2	default
	Learning Rate	0.1	0.1

Table 3.3: Parameter settings for different models across 5 and 10-fold cross validation for Reducing Health Expenditures

are experiencing a rise in prices with no increase in their salaries and are given low pension benefits, which could be a specific association related to reducing health expenditures. Other potentially significant predictors are rCSI, debt due to escalation security concerns because of armed violence and HH distance from Kakhovsky which appeared in four of the models. In addition to distance from Kakhovsky, other distance measures were recorded among the top important features such as distance from Skadovsky, Kalmiusky, Khersonskyi and many more. Moreover, food expenditures, long-term and short-term consumption expenditures are among the cash-related features that might effect the household's coping behavior. Lastly and most importantly, the displayed features suggest that the type of healthcare service and medicine the household desire are potentially associated to reducing health expenses.

3.3 Model Performance for Spending Savings Coping Mechanism using Downsampling

Dataset Dimension: $(3842, 177)$

Table 3.4: 5-fold stratified cross validated performance using different performance measures and confusion matrix values for Spent Savings

Type	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.67	0.66	0.68	0.6715	0.74	1314	655	1266	607
	XGB	0.70	0.71	0.70	0.7009	0.77	1321	549	1372	600
	RF	**0.73**	**0.76**	**0.71**	**0.7235**	**0.79**	**1320**	**461**	**1460**	**601**
	TabNet	0.70	0.72	0.67	0.6848	0.73	1246	536	1385	675
Tuned	Lasso	0.68	0.67	0.68	0.6795	0.74	1316	626	1295	605
	XGB	0.72	0.75	0.70	0.7163	0.78	**1318**	487	1434	**603**
	RF	**0.73**	**0.77**	**0.70**	**0.7209**	**0.79**	1298	**449**	**1472**	623
	TabNet	0.69	0.71	0.67	0.6767	0.71	1234	555	1366	687

Table 3.5: 10-fold stratified cross validated performance for Spent Savings

Method	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.67	0.66	0.68	0.677	0.74	1328	648	1273	593
	XGB	0.71	0.72	0.69	0.7022	0.77	1306	529	1392	615
	RF	**0.73**	**0.76**	**0.71**	**0.7230**	**0.79**	**1316**	**459**	**1462**	**605**
	TabNet	0.71	0.76	0.67	0.6892	0.72	1185	458	1463	736
Tuned	Lasso	0.68	0.67	0.68	0.6806	0.74	1324	630	1291	597
	XGB	0.73	0.74	0.71	0.7205	**0.79**	**1339**	492	1429	**582**
	RF	**0.75**	**0.78**	**0.71**	**0.7347**	**0.79**	1323	**421**	**1500**	598
	TabNet	0.72	0.76	0.68	0.6983	0.74	1231	465	1456	690

In the baseline models in table 3.5 for *spending savings*, RF emerges again as the strongest performer, achieving the highest $f_1Score = 0.73$, $recall = 0.76$, $accuracy = 72.30\%$, and $AUC = 0.79$. Compared to other models it also proves its robustness in distinguishing and predicting both classes. TabNet follows RF, with a strong $recall = 0.76$ but a lower $precision = 0.69$ which suggests more false positives compared to RF. Lasso falls behind with the lowest scores across most metrics, highlighting its limitations in this comparison.

After tuning, all models show improvement, with RF continuing to lead with an $f_1Score = 0.75$, $recall = 0.78$, and a 73.47% $accuracy$. XGB also shows significant improvements specifically scoring a 0.79 on the AUC (see figure 3.3), competing with RF. Lasso indicates slight improvements, yet it remains the least effective model. TabNet shows a strong $recall = 0.76$, but still scores low precision and accuracy compared to RF and XGB.

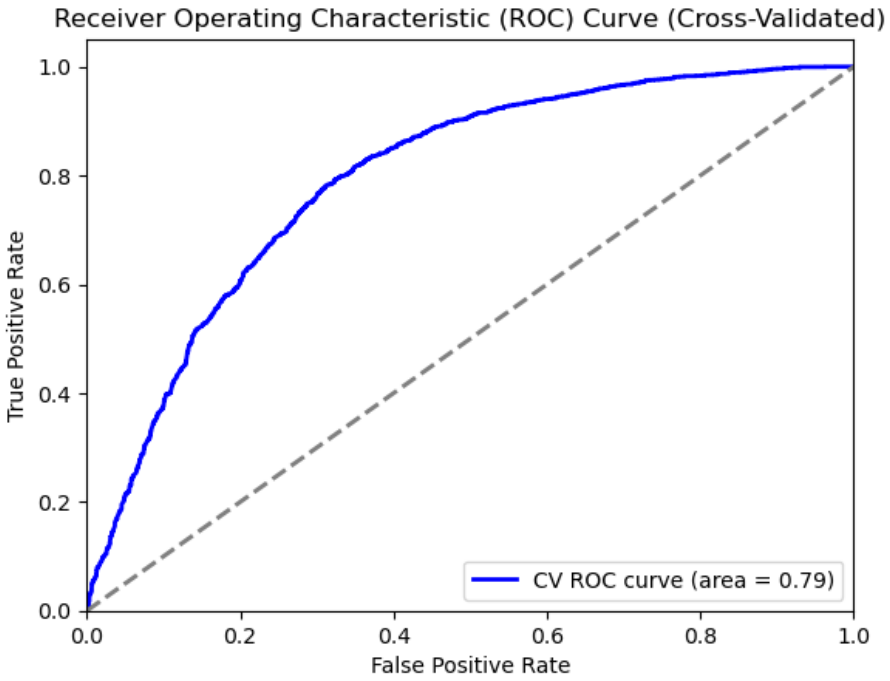


Figure 3.3: AUC-ROC curve of the Tuned Random Forest model with a 10 Fold Stratified CV

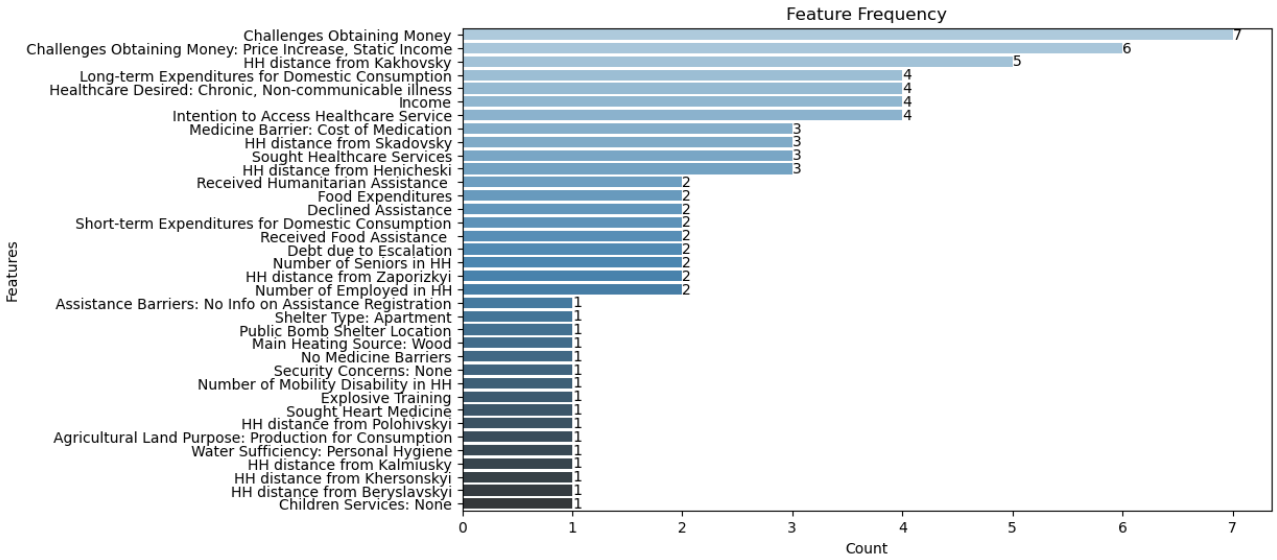


Figure 3.4: Optimal Set of Features: Spent Savings

3.4 Model Performance for Borrowing Food Coping Mechanism using Downsampling

Dataset Dimension: $(2286, 184)$

In the baseline comparison of models in table 3.7 for *borrowing food*, RF demon-

Table 3.6: 5-fold stratified cross validated performance using different performance measures and confusion matrix values for Borrowing Food

Type	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.78	0.78	**0.79**	0.7848	0.86	**904**	253	890	**239**
	XGB	0.78	0.79	0.78	0.7830	0.87	892	245	898	251
	RF	**0.80**	**0.83**	0.78	**0.7953**	**0.87**	869	**194**	**949**	274
	TabNet	**0.80**	0.81	0.78	0.7939	0.85	884	212	931	259
Tuned	Lasso	0.78	0.77	0.79	0.7852	0.86	910	258	885	233
	XGB	0.80	0.80	**0.80**	0**.8018**	**0.88**	**919**	229	914	**224**
	RF	**0.81**	**0.84**	0.78	**0.8014**	**0.88**	870	**181**	**962**	273
	TabNet	0.80	0.81	0.78	0.7939	0.84	887	215	928	256

Table 3.7: 10-fold stratified cross validated performance for Borrowing Food

Method	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.78	0.78	**0.79**	0.7826	0.86	**903**	257	886	**240**
	XGB	0.79	0.80	**0.79**	0.7896	**0.87**	894	232	911	249
	RF	**0.81**	**0.83**	0.78	**0.8**	**0.87**	877	**191**	**952**	266
	TabNet	0.80	0.81	**0.79**	**0.8**	0.85	**903**	217	926	**240**
Tuned	Lasso	0.78	0.78	0.79	0.7857	0.87	904	251	892	239
	XGB	0.80	0.80	**0.80**	0.8031	**0.88**	**917**	224	919	**226**
	RF	**0.81**	**0.84**	0.79	**0.8084**	**0.88**	884	**179**	**964**	259
	TabNet	0.81	0.83	0.79	0.8057	0.85	895	196	947	248

strates the strongest performance, achieving the highest $f_1Score = 0.81$, $recall = 0.83$, $accuracy = 80\%$, and $AUC = 0.87$. This indicates RF's strong overall performance in identifying true negatives. XGB also performs well, scoring $recall = 0.80$ and $AUC = 0.87$, and slightly passes RF in $f_1Score = 0.79$ and $precision = 0.79$. Lasso shows slightly lower performance across the metrics, with an $f_1Score = 0.78$ and $AUC = 0.86$. TabNet scored the lowest $AUC = 0.85$, while achieving a good balance between $recall = 0.81$ and$precision = 0.79$. After tuning, RF is in the lead with a stable $f_1Score = 0.81$, improved $recall = 0.84$, and the highest $accuracy = 80.84\%$ and $AUC = 0.88$ (see figure 3.5). XGB shows a slight improvement in $precision = 0.80$, $accuracy = 80.31\%$ and $AUC = 0.88$, making it a close competitor to RF. TabNet scores an 80.57% accuracy, $f_1 = 0.81$ and $recall = 0.83$ passing the XGB model. RF and TabNet are better at predicting the true negative classes than XGB. Lasso's performance remains consistent with minimal changes.

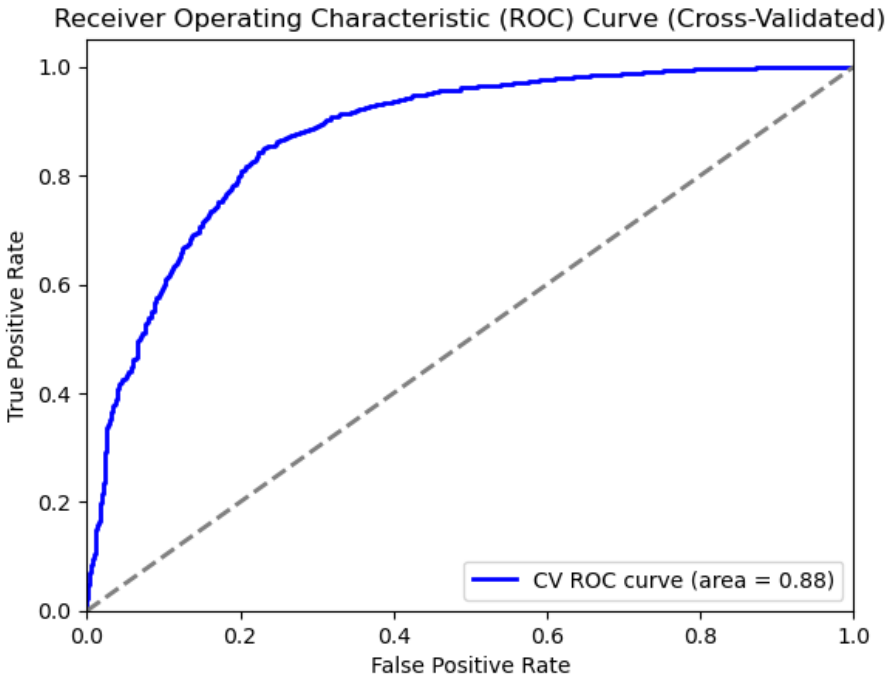


Figure 3.5: AUC-ROC curve of the Tuned Random Forest model with a 10 Fold Stratified CV

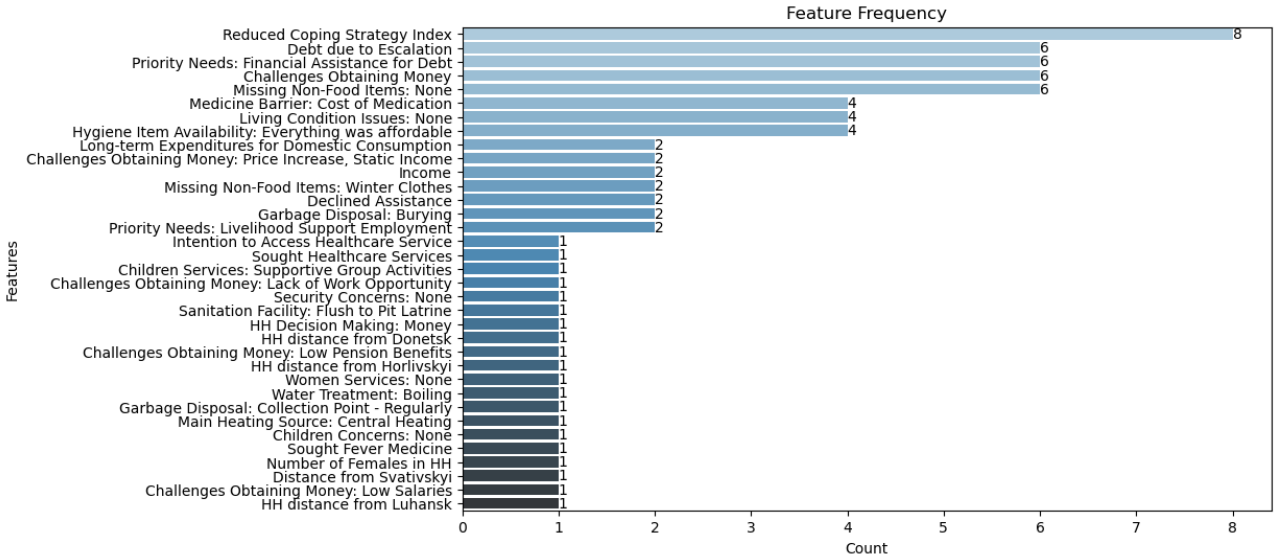


Figure 3.6: Optimal Set of Features: Borrow Food

3.5 Model Performance for Degrading Income Source Coping Mechanism using Downsampling

Dataset Dimension: $(834, 191)$

Table 3.8: 5-fold stratified cross validated performance using different performance measures and confusion matrix values for Degrading Income Source

Type	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.75	0.75	0.75	0.7494	0.82	313	105	312	104
	XGB	0.80	0.80	**0.80**	**0.7998**	0.86	**334**	84	333	**83**
	RF	**0.80**	**0.81**	0.79	0.7926	**0.87**	324	**80**	**337**	93
	TabNet	0.79	0.82	0.77	0.7841	0.82	314	77	340	103
Tuned	Lasso	0.76	0.77	0.75	0.7578	0.84	312	97	320	105
	XGB	0.81	0.81	**0.81**	0.8058	**0.88**	**336**	81	336	**81**
	RF	**0.82**	**0.83**	0.80	**0.8118**	0.87	331	**71**	**346**	86
	TabNet	0.80	0.79	0.80	0.7997	0.84	337	87	330	80

Table 3.9: 10-fold stratified cross validated performance for Degrading Income Source

Method	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.74	0.74	0.75	0.7470	0.83	315	109	308	102
	XGB	0.79	0.79	0.78	0.7853	0.86	**327**	89	328	**90**
	RF	**0.80**	**0.83**	0.78	0.7973	**0.87**	317	**69**	**348**	100
	TabNet	0.81	**0.83**	**0.79**	**0.8034**	0.85	323	70	347	94
Tuned	Lasso	0.77	0.76	0.78	0.7710	0.85	326	100	317	91
	XGB	0.80	0.81	0.79	0.7962	0.88	327	80	337	90
	RF	0.81	**0.85**	0.78	**0.8045**	**0.88**	317	**63**	**354**	100
	TabNet	**0.82**	0.83	**0.82**	**0.8202**	0.84	**339**	72	345	**78**

In the baseline comparison in table 3.9 for *degrading income source*, RF stands out again with the highest $AUC = 0.87$ and scored $f_1Score = 0.80$, $recall = 0.83$, and $accuracy = 0.7973$. While TabNet score better $Recall$, F_1 and $Precision$, it scores a low $AUC = 0.85$. XGB also performs well with an $AUC = 0.86$, but is slightly behind in $precision$, $recall$ and f_1. Lasso shows the lowest metrics compared to all models.

After tuning, all models improve, with TabNet in the lead achieving the highest $f_1Score = 0.82$, $precision = 0.82$ and accuracy of 82.02%. RF has highest $recall = 0.85$ and $AUC = 0.88$ (see figure 3.7) and improved Accuracy of 80.45%. XGB also scores and AUC of 0.88 but scores lower $precision$, $recall$ and F_1. Lasso's performance remains consistent with minimal changes.

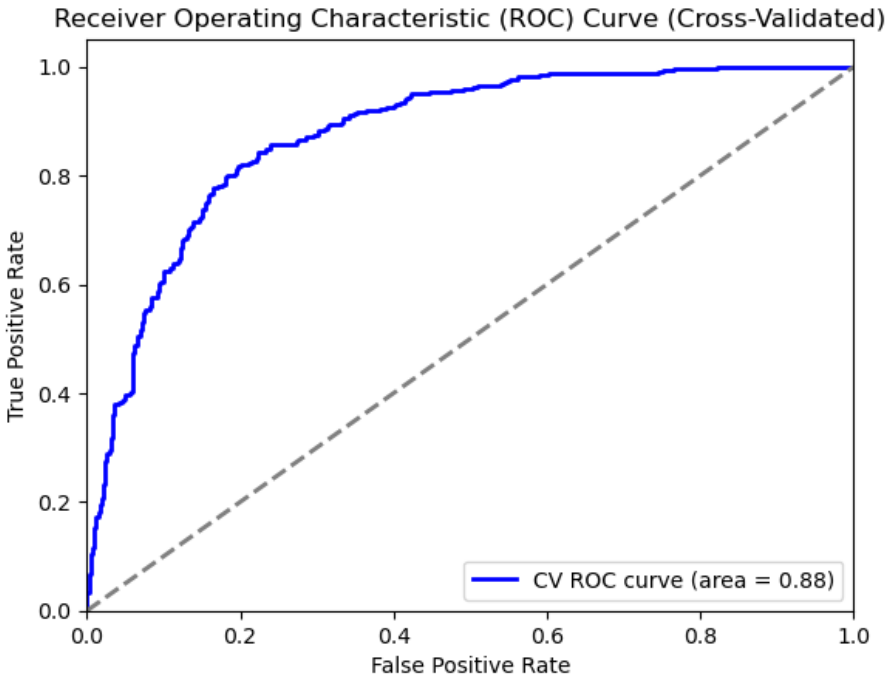


Figure 3.7: AUC-ROC curve of the Tuned Random Forest model with a 10 Fold Stratified CV

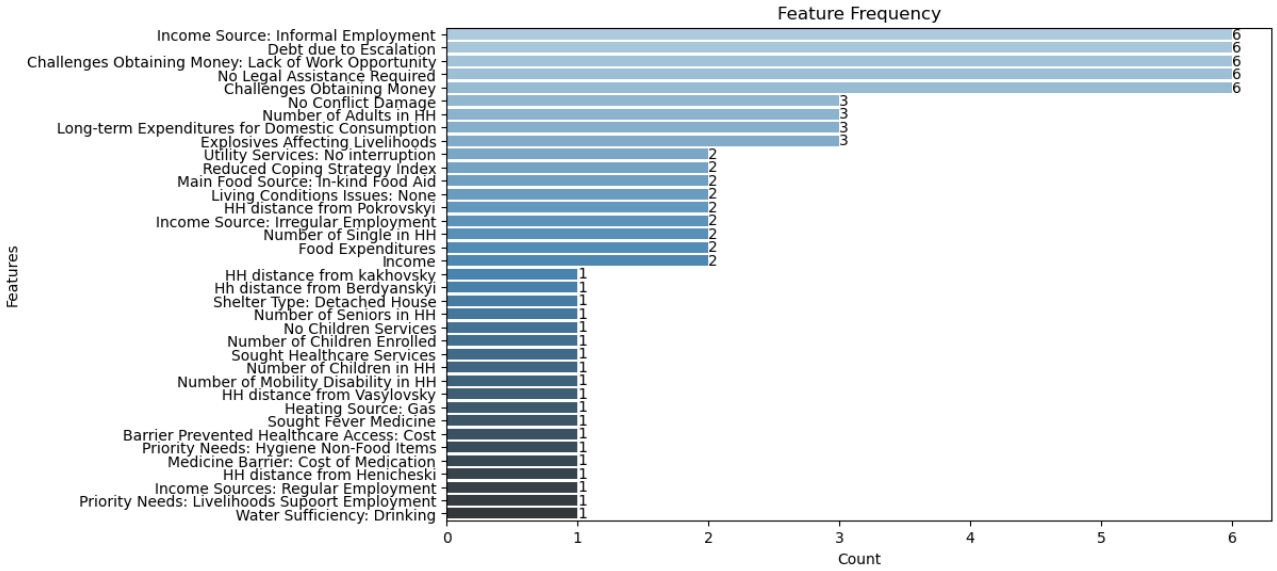


Figure 3.8: Optimal Set of Features: Degrade Income Source

3.6 Model Performance for Reduce Education Expenditures Coping Mechanism using Downsampling

Dataset Dimension: $(678, 191)$

Table 3.10: 5-fold stratified cross validated performance using different performance measures and confusion matrix values for Reducing Education Expenditures

Type	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.70	0.69	0.72	0.7079	0.79	247	106	233	92
	XGB	0.74	0.75	0.74	0.7433	0.81	250	85	254	89
	RF	**0.77**	**0.78**	**0.76**	**0.7684**	**0.84**	**255**	**73**	**266**	**84**
	TabNet	0.76	**0.78**	0.75	0.7595	0.81	251	75	264	88
Tuned	Lasso	0.73	0.71	0.75	0.7330	0.81	258	100	239	81
	XGB	0.75	0.75	0.76	0.7566	0.84	259	85	254	80
	RF	**0.78**	**0.77**	0.78	0.7758	**0.85**	264	**77**	**262**	75
	TabNet	0.77	0.75	**0.80**	**0.7802**	0.81	**275**	85	254	**64**

Table 3.11: 10-fold stratified cross validated performance for Reducing Education Expenditures

Method	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.72	0.71	0.72	0.7182	0.79	247	99	240	92
	XGB	0.74	0.74	0.73	0.7345	0.81	246	87	252	93
	RF	0.77	0.76	0.78	0.7713	**0.84**	265	81	258	74
	TabNet	**0.79**	**0.78**	**0.81**	**0.7964**	0.83	**275**	**74**	**265**	**64**
Tuned	Lasso	0.73	0.72	0.74	0.7374	0.81	255	94	245	84
	XGB	0.75	0.74	0.75	0.7507	0.83	257	87	252	82
	RF	0.79	0.80	0.78	0.7861	**0.85**	262	68	271	77
	TabNet	**0.80**	**0.81**	**0.80**	**0.8023**	**0.84**	**269**	**64**	**257**	**70**

After tuning, in table 3.11 for *reducing education expenditures*, TabNet remains the top model with the highest $f_1Score = 0.80$, $recall = 0.81$, and $accuracy = 80.23\%$. The $recall = 0.80$ and $accuracy = 78.61\%$ of the RF model improved and it scored the highest AUC score 0.85 (see figure 3.9), closely matching TabNet in performance. XGB and Lasso improve slightly but they remain behind the top models.

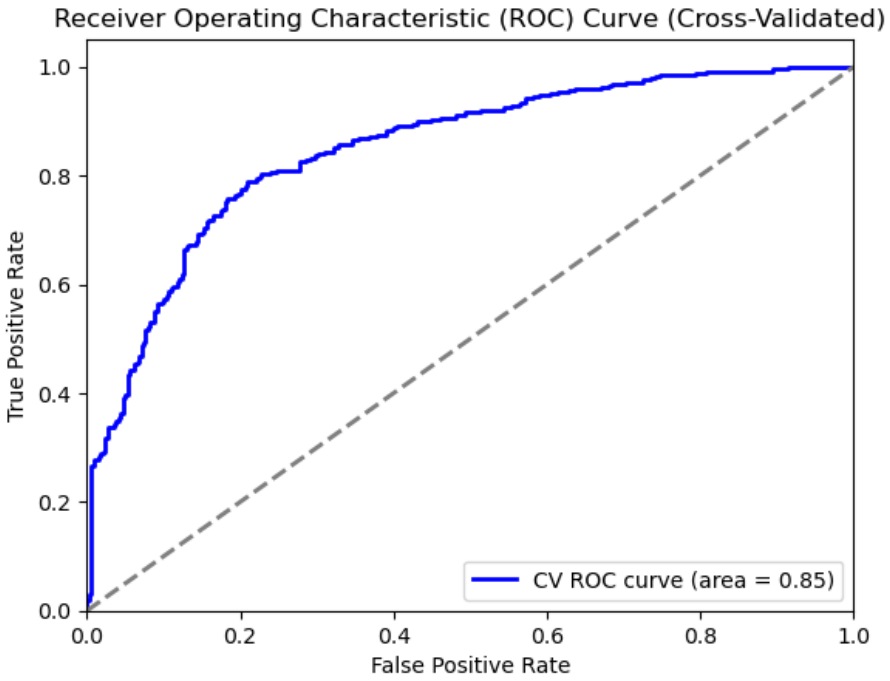


Figure 3.9: AUC-ROC curve of the Tuned Random Forest model with a 10 Fold Stratified CV

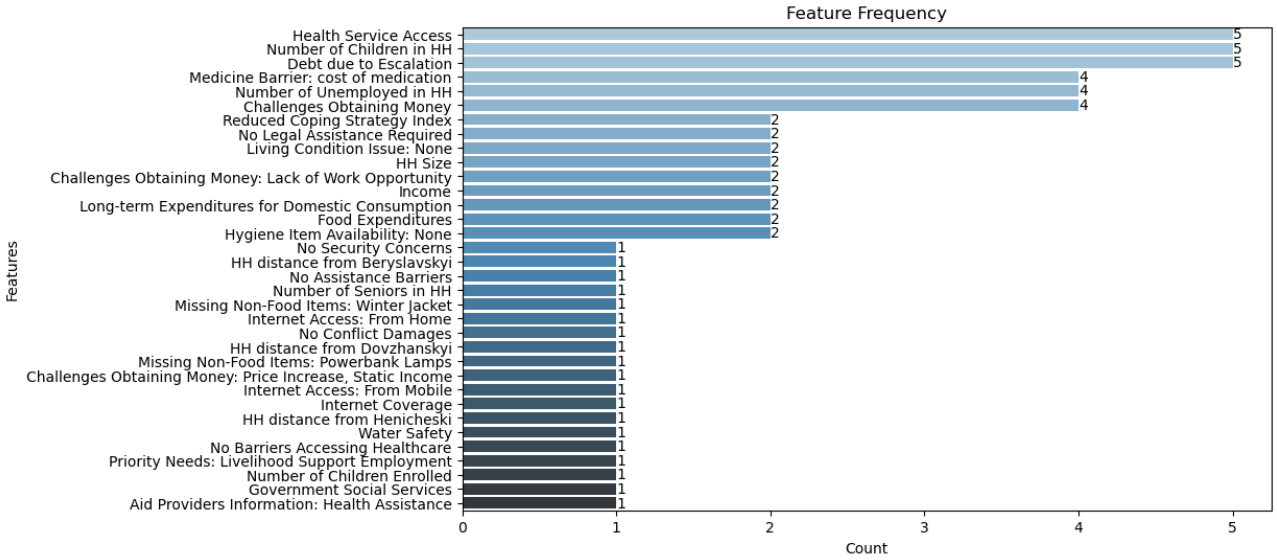


Figure 3.10: Optimal Set of Features: Reduce Education Expenditures

3.7 Model Performance for Sell Household Assets Coping Mechanism using Downsampling

Dataset Dimension: $(656, 194)$

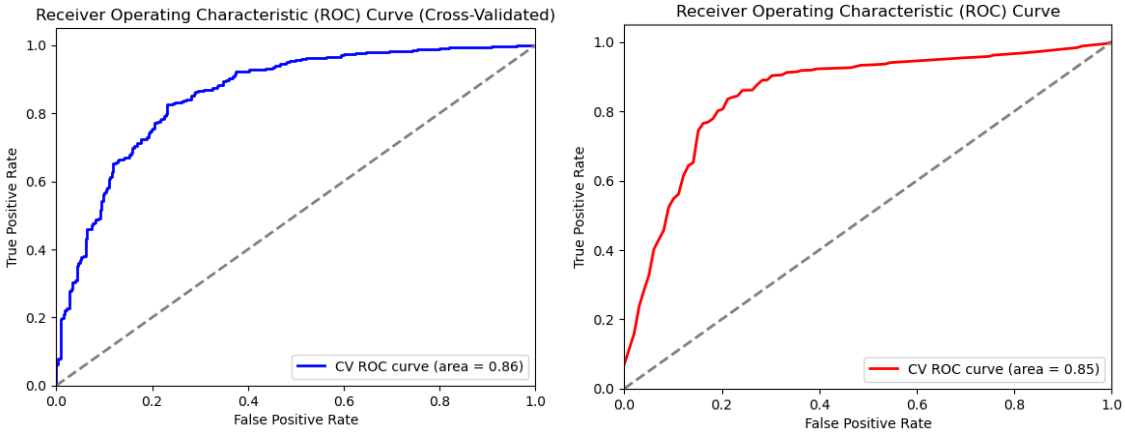
Table 3.12: 5-fold stratified cross validated performance using different performance measures and confusion matrix values for Selling HH Assets

Type	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.74	0.72	0.75	0.7423	0.81	250	91	237	78
	XGB	0.76	0.78	0.75	0.7576	0.83	242	73	255	86
	RF	**0.79**	**0.80**	0.77	0.7820	**0.85**	251	66	262	77
	TabNet	**0.79**	**0.80**	**0.79**	**0.7896**	0.82	**275**	**67**	**261**	**71**
Tuned	Lasso	0.77	0.77	0.77	0.7682	0.84	252	76	252	76
	XGB	0.79	0.80	0.78	0.7850	0.85	253	66	262	75
	RF	0.80	0.81	0.78	0.7911	**0.86**	252	**61**	**267**	76
	TabNet	**0.81**	**0.81**	**0.80**	**0.8048**	0.84	**263**	63	265	**65**

Table 3.13: 10-fold stratified cross validated performance for Selling HH Assets

Method	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.75	0.74	0.76	0.7530	0.82	251	85	243	77
	XGB	0.76	0.77	0.75	0.7545	0.83	244	77	251	84
	RF	0.80	0.82	0.77	0.7896	**0.86**	248	58	270	80
	TabNet	**0.82**	**0.84**	**0.80**	**0.8185**	0.84	**261**	**52**	**276**	**67**
Tuned	Lasso	0.77	0.77	0.77	0.7713	0.84	252	74	245	76
	XGB	0.79	0.80	0.78	0.7865	0.85	255	67	261	73
	RF	0.80	0.83	0.77	0.7926	**0.86**	249	57	271	79
	TabNet	**0.83**	**0.84**	**0.82**	**0.8262**	**0.85**	**267**	**53**	**275**	**61**

The tuned models in table 3.13 for *selling household assets*, show that TabNet further strengthens its lead, achieving the highest $f_1 Score = 0.83$, $recall = 0.84$, and $accuracy = 82.62\%$ with an $AUC = 0.85$. RF continues to perform well, maintaining a the highest $AUC = 0.86$ (see figure 3.11). XGB shows improvement across all metrics but still retained it position behind TabNet and RF. Lasso remains the least competitive with an *accuracy* of 77.13% .



(a) (b)

Figure 3.11: AUC-ROC curve of the Tuned Random Forest (a) and Tuned TabNet model (b) with a 10 Fold Stratified CV

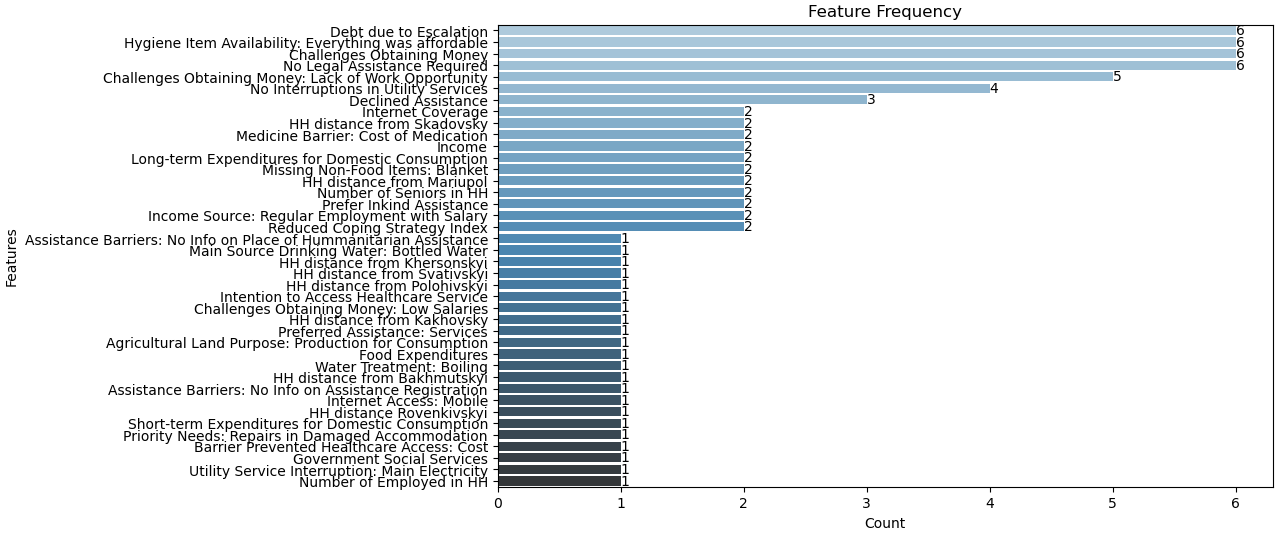


Figure 3.12: Optimal Set of Features: Sell household Assets

3.8 Model Performance for Eat Elsewhere Coping Mechanism using Downsampling

Dataset Dimension: $(458, 197)$

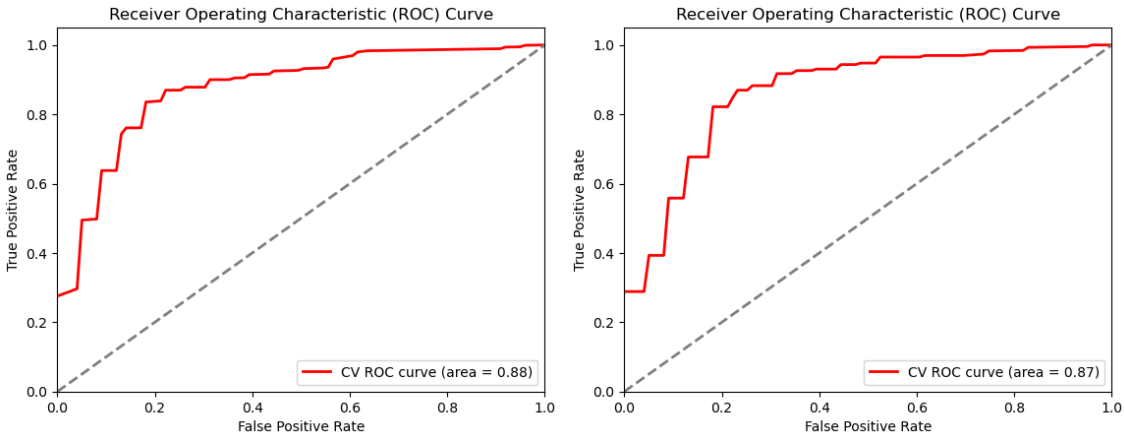
Table 3.14: 5-fold stratified cross validated performance using different performance measures and confusion matrix values for Eating Elsewhere

Type	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.75	0.75	0.75	0.7532	0.83	173	57	172	56
	XGB	0.75	0.76	0.74	0.7489	0.84	168	54	175	61
	RF	0.79	0.81	0.77	0.7838	0.85	173	43	186	56
	TabNet	**0.82**	**0.86**	**0.79**	**0.8166**	**0.85**	**176**	**31**	**198**	**53**
Tuned	Lasso	0.79	0.79	0.78	0.7838	0.86	178	48	181	51
	XGB	0.76	0.76	0.75	0.7576	0.85	172	54	175	57
	RF	0.79	**0.83**	0.76	0.7816	**0.86**	169	**40**	**189**	60
	TabNet	**0.81**	0.81	**0.80**	**0.8034**	**0.86**	**182**	43	186	**47**

Table 3.15: 10-fold stratified cross validated performance for Eating Elsewhere

Method	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.74	0.74	0.75	0.7445	0.82	171	59	170	58
	XGB	0.77	0.77	0.76	0.764	0.84	174	53	176	55
	RF	0.80	0.82	0.78	0.7947	0.86	176	41	188	53
	TabNet	**0.85**	**0.86**	**0.84**	**0.8493**	**0.88**	**192**	**32**	**197**	**37**
Tuned	Lasso	0.79	0.80	0.78	0.7882	0.87	178	46	183	51
	XGB	0.77	0.78	0.76	0.7641	0.85	172	51	178	57
	RF	0.81	0.82	0.79	0.8034	0.87	180	41	188	49
	TabNet	**0.85**	**0.89**	**0.82**	**0.8451**	**0.87**	**184**	**26**	**203**	**45**

In table 3.15 for *eating elsewhere*, the tuned TabNet model maintains its lead, with highest $recall = 0.89$, $f_1Score = 0.85$, and $accuracy = 84.51\%$. RF improves with an 80.34% *accuracy* and 0.81 f_1Score, remaining a strong competitor with an AUC of 0.87 (see figure 3.13). Lasso shows significant improvement in r_1Score (0.79) and *recall* (0.80) scoring a high AUC (0.87).



(a) (b)

Figure 3.13: AUC-ROC curve of the Baseline TabNet (a) and Tuned TabNet model (b) with a 10 Fold Stratified CV

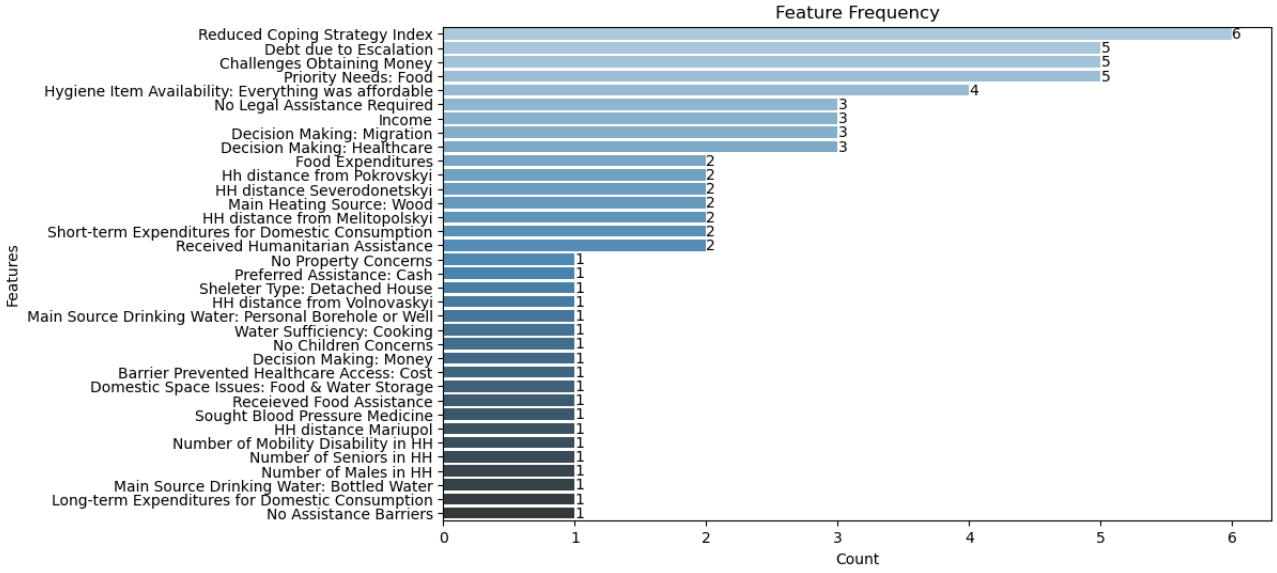


Figure 3.14: Optimal Set of Features: Eat Elsewhere

3.9 Model Performance for Ask Stranger Coping Mechanism using Downsampling

Dataset Dimension: $(342, 196)$

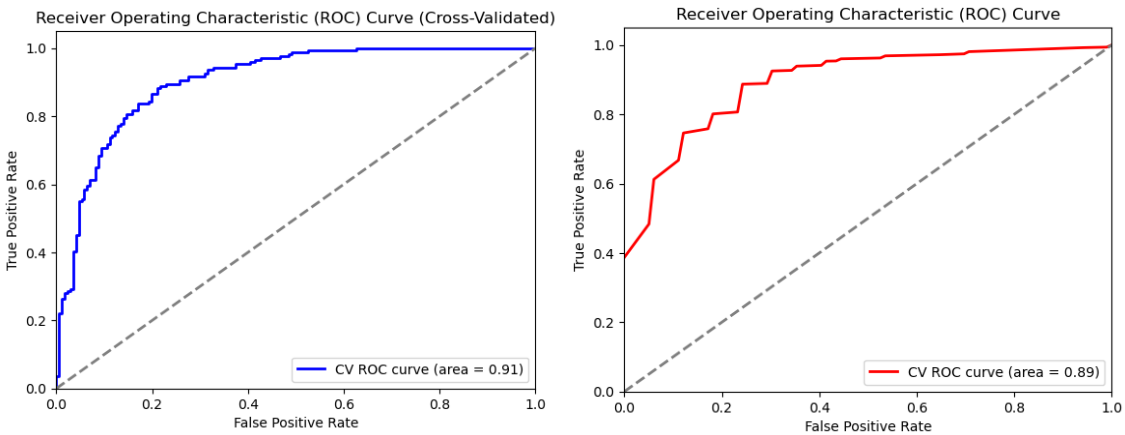
Table 3.16: S5-fold stratified cross validated performance using different performance measures and confusion matrix values for Ask Stranger

Type	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.77	0.76	0.78	0.7719	0.86	134	41	130	37
	XGB	0.80	0.81	0.79	0.7953	0.88	133	32	139	38
	RF	0.82	**0.84**	0.81	0.8187	**0.89**	137	**28**	**143**	34
	TabNet	**0.83**	0.81	**0.85**	**0.8305**	**0.89**	**146**	33	138	**25**
Tuned	Lasso	0.81	0.82	0.81	0.8128	0.89	138	33	140	33
	XGB	0.81	0.81	0.82	0.8157	**0.90**	141	33	138	30
	RF	0.82	0.84	0.80	0.8157	**0.90**	135	27	144	36
	TabNet	**0.87**	**0.87**	**0.88**	**0.8713**	**0.89**	**150**	**23**	**148**	**21**

Table 3.17: 10-fold stratified cross validated performance for Ask Stranger

Method	Model	F_1	Recall	Precision	Accuracy	AUC	TP	FN	TN	FP
Baseline	Lasso	0.78	0.78	0.77	0.7748	0.85	132	38	133	39
	XGB	0.80	0.80	0.79	0.7953	0.88	135	34	137	36
	RF	0.81	**0.82**	0.80	0.8099	**0.89**	137	**31**	**140**	34
	TabNet	**0.84**	0.81	**0.88**	**0.8482**	0.87	**152**	33	138	**19**
Tuned	Lasso	0.80	0.80	0.80	0.8011	0.88	137	34	137	34
	XGB	0.83	0.82	0.83	0.8274	**0.91**	142	30	141	29
	RF	0.82	0.81	0.83	0.8245	**0.90**	143	32	139	28
	TabNet	**0.88**	**0.88**	**0.87**	**0.8773**	**0.89**	**149**	**20**	**151**	**22**

Lastly, after tuning, table 3.17 for *asking strangers*, shows improvements at the level of all the models, with TabNet still leading with an f_1Score of 0.88, *recall* of 0.88, *precision* of 0.87, and *accuracy* of 87.73%. XGB marks a notable increase in f_1Score (0.83) and *precision* (0.83). XGB model records the highest *AUC* score 0.91 (see figure 3.15).



(a) (b)

Figure 3.15: AUC-ROC curve of the Tuned XGBoost (a) and Tuned TabNet model (b) with a 10 Fold Stratified CV

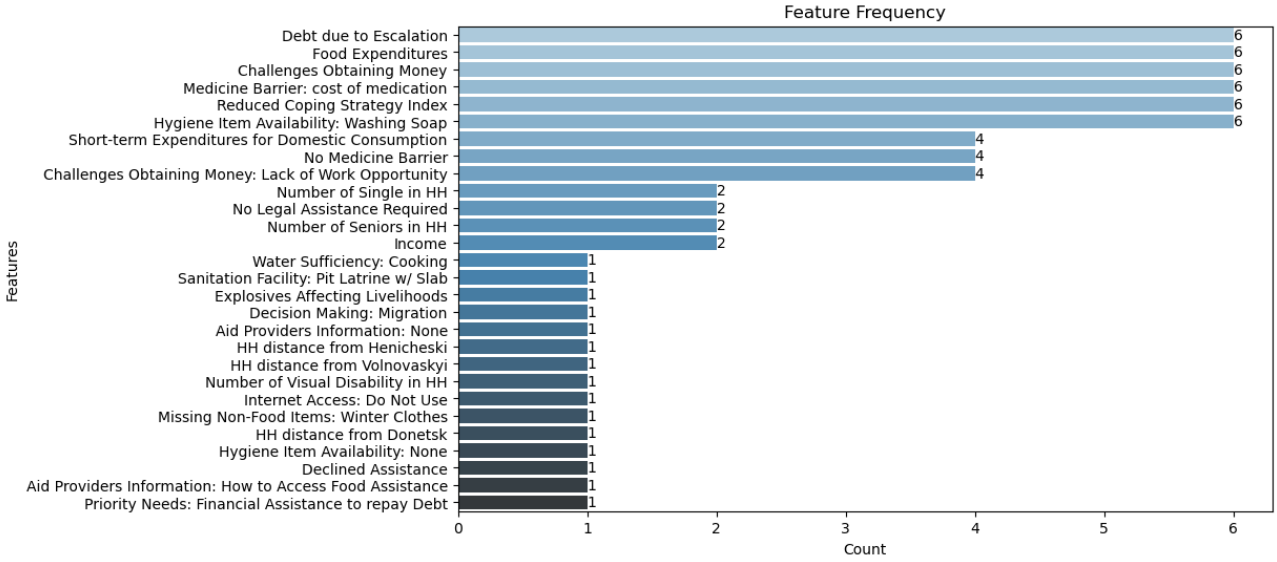


Figure 3.16: Optimal Set of Features: Ask Stranger

3.10 Lasso Model used in Quantifying the Associations

Target variable	Precision	Recall	F1	AUC	Cost
Reduce Health Expenditures	0.38	0.82	0.51	0.86	0.1
Spent Savings	0.26	0.72	0.38	0.76	0.01
Borrow Food	0.28	0.80	0.41	0.86	0.1
Degrade Income Source	0.11	0.74	0.19	0.85	0.1
Reduce Education Expenditures	0.08	0.79	0.15	0.85	0.1
Sell household Assets	0.08	0.70	0.14	0.81	0.1
Eat Elsewhere	0.07	0.74	0.14	0.89	0.1
Ask Stranger	0.05	0.65	0.09	0.84	0.1

Table 3.18: Lasso model metrics for all the negative coping mechanism

3.10.1 *Lasso Summary Table*

Table 3.18 displays the model metrics of a Lasso model for each of the target variables included in the study. Setting the class weight parameter as *'balanced'* improved the *recall* score, however it affected the *precision* of the model. The model now is falsely classifying class 0 predictions as class 1 more, which we call a type 1 error. This type of error is not harmful in our case since classifying a household as does not employ a negative coping mechanism as household that cope, is considered a precautionary measure. As we aim to target households that employ negative coping mechanisms, our concern focuses on not committing a type 2 error. In other words, we prioritize a model that correctly classifies class 1 and minimizes the cases where households actually employing a negative coping mechanism are classified as not employing. All models score a high *AUC* score ranging between 0.81 and 0.89 except for the *Spent Savings* model which scored the lowest *AUC* score of 0.75. A high *AUC* score is a good indicator that the model is able to distinguish between class $'0'$ and class $'1'$. The learning curve for *reducing health expenditures* outcome displayed in figure 3.17, showed a good fit. Both the training and validation score were increasing, close to each other and stabilizing as the number of training examples increase. Also, figure C.2 (c), referring to the *Borrowed Food* outcome, showed a good fit. The rest of the learning curves show overfitting. Specifically, figures C.3 (e) and C.4 (g) indicate mild overfitting where the training score was higher than the validation score but the gap between them was small. The validation score increases as more data was added. Lastly, figures C.1 (b), C.2 (d), C.3 (f) and C.4 (h) indicate significant overfitting where the gap between the training and cross validation score was large suggesting that the model performs poorly on unseen data. The main reason behind this was possibly due to insufficient minority class data.

3.10.2 *Optimal Feature Set and Lasso Associations*

To further assess whether the features illustrated in graph 3.2 hold a predictive power over the reducing health expenses, we used Lasso. We explored the association and excluded features that might have ranked in top 10 features, but did not display actual importance in the model. Table 3.19 presents the final model's coefficients along with their significant p-values. The coefficients of the significant features are exponentiated to represent the odds of an outcome given a one unit increase in the feature. The odds of reducing health expenditures among household who face challenges obtaining money is $e^{1.1378} \approx 3.11$ times the odds of coping for those who do not face challenges obtaining money accounting for all the other features in the model. Also, households who deal with security concerns due to armed violence are more likely to reduce health expenditures by 1.7 times compared to households that are not concerned. In the context of healthcare, going from not experiencing med barriers to experiencing medication barriers due to medication cost, the odds of reducing health expenditures increases multiplicatively by 1.52. This adds on to the previously mentioned literature that also medicine cost barriers are directly associated with coping by reducing health expenses. Furthermore, going from not having debt to being in debt due to escalation, the odds of coping by reducing health expenditures increases multiplicatively by $e^{0.4586} \approx 1.58$ accounting for all the other features in the model. For one Ukrainian Hryvnia increase in the income, the odd of reducing health expenditures decreases multiplicatively by $e^{0.18} \approx 1.19$ holding all other features in the model constant. Indicating that the income is a protective factor.

3.10.3 *Reduce Health Expenditures Lasso Model Coefficients*

	Coefficient	Std. Error	Z-Score	P-Value
Intercept	-0.249800	0.098200	-2.542600	0.011000
Medicine Barrier: Cost of Medication	0.420900	0.010100	41.575100	0.000000
HH distance from Skadovsky	0.127300	0.009800	12.961800	0.000000
Income	-0.180400	0.011300	-15.921700	0.000000
Food Expenditures	-0.037700	0.011100	-3.385200	0.000700
Healthcare Service Desired: Chronic, Non-communicable illness	0.231800	0.010000	23.116900	0.000000
Healthcare Service Desired: Diagnostic	0.118500	0.009400	12.595600	0.000000
Challenges Obtaining Money	1.137800	0.028300	40.189400	0.000000
Reduced Coping Strategy Index 1	0.365400	0.022400	16.334200	0.000000
Reduced Coping Strategy Index 2	0.489500	0.056800	8.612200	0.000000
Security Concerns: Armed violence/Shelling	0.533900	0.030800	17.346700	0.000000
Debt Due to Escalation	0.458600	0.027300	16.815000	0.000000
Hygiene Item Availability: Everything was affordable	-0.218900	0.024400	-8.979000	0.000000
Priority Needs: Drinking Water	-0.115700	0.029600	-3.910800	0.000100
Challenges Obtaining Money: Price Increase, Static Income	0.162300	0.028600	5.679100	0.000000
Declined Assistance	-0.560800	0.023900	-23.427200	0.000000
Main Food Source: Inkind Food Aid	0.360800	0.042400	8.507800	0.000000
Received Humanitarian Assistance	0.305700	0.030000	10.192800	0.000000
Women Support Services: Hotline for Violence	0.080600	0.028300	2.845300	0.004400
Agricultural Land Purpose: Renting	0.105600	0.027700	3.819700	0.000100
No Concerns for Children	-0.281200	0.022000	-12.787200	0.000000
Food Consumption Score 1	0.000000	0.080200	0.000000	1.000000
Food Consumption Score 2	-0.326800	0.073700	-4.431400	0.000000
Water Sufficiency: Cooking	-0.357400	0.063500	-5.632100	0.000000
Explosives Effecting Livelihoods	0.467600	0.036500	12.820600	0.000000
Internet Access: Mobile	-0.147800	0.022000	-6.721100	0.000000
Sanitation Facility: Flush to Pit Latrine	-0.238700	0.020800	-11.487900	0.000000
Income Source: Pension	0.063600	0.020600	3.087300	0.002000

Table 3.19: Reduce Health Expenditure Lasso Model Summary

Moving on to the *Spent Savings* outcome displayed in figure 3.4, we notice challenges obtaining money ranks again as the most prominent feature for spending savings. One of the top features identified was distance from Kakhovsky which appeared in five of the models. Income, debt and health-related features such as

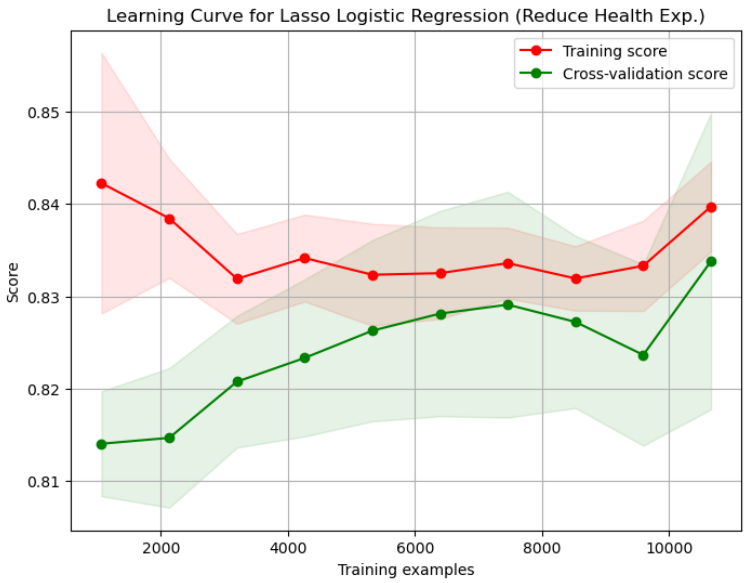


Figure 3.17: Reduce Health Expenditure Learning Curve of Lasso

access to healthcare services and medicines make a significant appearance. Household characteristics, such as number of seniors, number of disabaled and number of employed individuals in household are captured indicating a possible effect on households coping through spending savings. It could be due to unemployment, low wages, buying medications for disability, healthcare visits for seniors and so on. Other expenditures that appear within the prominent features are long-term and short-term expenditures for domestic consumption. Moreover food expenditures along side receiving food assistance and income could possibly play a major role in deciding whether the household copes or not given that food is one of the main living necessities. Finally, multiple proximity to frontline features are captured such as Skadovsky, Zaporizkyi, Svativskyi and Kalmiuskyi. Barriers to accessing medicine, healthcare services and assistance, as well as the type of assistance received, could significantly impact the odds of spending savings. To quantify these associations and interpret their effect on the outcome, we used Lasso. The final model C.1 suggests that the odds of spending savings among household who face challenges obtaining money is $e^{0.3726} \approx 1.45$ times the odds of coping for those who do not face challenges obtaining money accounting for all the other features in the model. As for the long-term expenditures, as they increase in one Ukrainian Hryvnia, the odds of spending savings increases by $e^{0.3} \approx 1.35$. Kalmiusky, located in Donetsk oblast, is positively associated with spending savings indicating that it is a risk factor to be distant from the frontline. On the other hand, number of employed and number of seniors in household are negatively associated with spending savings. This means that as the number of seniors (employed individuals) in a household increase by one senior (employed) member, the odds of spending savings decreases multiplicatively by $e^{0.104} \approx 1.11$. The odds that a household spent savings among those that are in debt due to escalation, is $e^{0.341} \approx 1.4$ times the odds to spend savings when

household are not in debt due to escalation, accounting for the other features in the model. Finally, the odds of coping among households intending to access healthcare due to experiencing a medical problem is 1.26 times those who cope but do not intend to access healthcare services.

Our findings in figure 3.6 for *borrowing food* indicate a significant ranking for the reduced coping strategy index where it appeared as a top feature across all the models. Debt and the need for financial assistance to repay debt were very prominent as they are present in six models. Again, challenges obtaining money has a striking presence within the frequently appeared features between the models. A compelling appearance of type of challenges in obtaining money is observed in the graph representation including lack of work opportunity, low pension benefits, low salaries and price increase with static incomes. All these scenarios might be impacting the coping style of a household specifically by borrowing food. Furthermore, income, medication cost barriers, and long-term expenditures play a role in the borrowing food coping strategy. Proximity to Donetsk, Luhansk, and Svativskyi showcase a subtle presence as they were detected as possibly prominent by one of the models only. Lastly and for the first time, the results suggest that number of females in a household plays a role in whether the household borrows food or not. Gender can significantly affect the borrowing behavior of a household due to household roles, decision-making power, and access to resources [12], [24]. From table C.2, we discovered that not only is challenges in obtaining money a risk factor of borrowing food, but also the lack of work opportunity and low salaries are a key factors in defining the household's borrowing behavior. That is, the odds of borrowing food for those who lack work opportunities is $e^{0.22} \approx 1.246$ times the odds of household borrowing food for those who have work opportunities. One of the most essential feature to tackle here is the rCSI. The rCSI is a categorical variable consisting of 3 categories: relative (0), moderate (1), severe (2) food insecurity. Given that relative food insecurity (0) is our reference category and accounting for the other features in the model, going from relative to moderate food insecurity, the odds of borrowing food increases multiplicatively by $e^{0.6939} \approx 2$. Similarly, going from relative to severe food insecurity, the odds of borrowing food increases multiplicatively by $e^{1.44} \approx 4.2$. Furthermore, the odds that a household borrowing food among those that are in debt due to escalation, is $e^{1.26} \approx 3.52$ times the odds to borrow food when household are not in debt due to escalation, accounting for the other features in the model. The model highlights the households' needs of financial assistance to pay debt. We found that the odds of borrowing food among those who need financial assistance is $e^{0.8165} \approx 2.26$ times the odds of those who borrow food but do not need financial assistance to repay debt.

According to graph 3.8 for *degrading income source*, households who rely on informal employment for income and those which face challenges in obtaining money, due to lack of work opportunities, were prominent in six models. We also notice that there might be an association between degrading income source and some household characteristics like number of adults, number of single members, number of children in household and number of children enrolled in school. A significant number of proximity to frontline features appear prominent but only in one of the model. These

features include distances from Kakhovsky, Berdyanskyi and Henicheski. Other possible indicators of degrading income source can be the healthcare barriers such as cost or accessing the services. The Lasso model in table C.3 backs up our previous suppositions. The odds of employing the negative coping strategy for those who face challenges acquiring money due to lack of work opportunities is 1.96 times those who do not face these challenges. Also, as the number food expenses and long-term expenditures increase by one Ukrainian Hryvnia, the odds of degrading income source increases by 1.18 and 1.145 respectively indicating that increases expenses is a risk factor for such coping mechanism. Moreover, household income acquired from informal employment puts the household at risk for degrading income source. On the other hand, some protective factors from degrading income source are household which have no living condition issues, no conflict damage, no interruption in utility services and do no require legal assistance.

The best features that are potential precursors of *reducing educational expenses* are presented in figure 3.10. The most frequent features were accessing health services, number of children in household and debt due to escalation which appeared in five out of the eight models. Then, they are followed by other household characteristics such as number of members in household and number of children enrolled in school. Additionally, food expenditures and long-term expenditures for domestic consumption indicate a relevance to the outcome by appearing in two models. Conversely, while proximity from Beryslavskyi, Dovzhanskyi and Henicheski appear less frequently, their association with the outcome might still be significant. Although the previous results suggest that the household size is a protective factor, evidence from the Lasso model in table C.4 suggests that it is positively associated with reducing education expenditures. Accounting for the features in the model, as the number of members increase one, the odds of reducing education expenditures increases multiplicatively by 1.23, making is a risk factor. Once again, the feature challenges obtaining money showcases a consistent positive association with the outcome. Results show that the odds of reducing education expenditures among those who do not face such challenges versus those who do, increases multiplicatively by 2.3. Additionally, proximity to Beryslavskyi was found to be positively associated with reducing education expenditures meaning that the closer the household is to this raion, the more likely to use this negative coping mechanism. Lastly, going from household that has security and safety concerns to household that do not experience such concerns, the odds of reducing education expenses decrease multiplicatively by 1.9. This means that the household that is concerned about it's members safety, is likely to resort to drop the children from school or resort to different means of learning hence resulting in the reduction of education expenditures.

In the bar graph 3.12 for *selling household assets*, the most frequent features the might be potential predictors of selling household assets are debt due to escalation and challenges obtaining money where they appear in six out of eight models. More specifically, lack of work opportunities and low wages are what might cause households to cope by selling household assets. Medical cost and assistance barriers consistently show up as potential predictors for coping as well as several expenditure indicators. Multiple distance from frontline features suggest some association to the

coping mechanism. In the findings of the Lasso model in table C.5, the factors that protect the household from selling their assets are the number of seniors in the house, affording all hygiene items, requiring no legal assistance, facing no interruptions in utility services and depending on a salary from regular employment. The results indicate that lack of work opportunities and low wages drive households to sell their assets to secure some cash to meet their immediate financial needs. Consistently, the results suggest that being in debt due to escalation and facing challenges acquiring money are risk factors of the outcome. The odds of selling household assets for household in debt due to escalation is 4 times the odds of selling household assets for household that are not in debt, accounting for all other features. Furthermore, households going from relative to severe food insecurity, the odds of selling household assets increases multiplicatively by 2.28. This indicates that households suffering from severe food insecurity tend to sell household assets to cope with the circumstances. This behavior emerges from the urgent need to generate immediate cash to meet basic needs, such as purchasing food or paying for essential services.

From bar graph 3.14 the most frequent features that are potentially related to *eating elsewhere* are debt due to escalation, challenges obtaining money, rCSI and priority need for food. These four features are a key factor in assessing whether a household eats elsewhere or not and they appear in at least five out of the eight models. Multiple proximity to frontline features are significant in the models where these could give us some insight on the association of the outcome with respect to the geographical location of a household. Furthermore, gender, specifically the number of males in the house, seems to play an important role for eating elsewhere. We will discover whether there is an association between them from the Lasso model summary. From C.6, we find that households going from relative to severe food insecurity, the odds of eating elsewhere increases multiplicatively by 6.45. Interestingly, as the number of males in the house increases by one, the odds of eating elsewhere increases by 1.26. This could be due safety concerns or economic pressures that drive them to seek food outside the house. Moreover, the farthest the household is from Melitopolskyi, situated in Zaporizhzhia Oblast, the more likely to eat elsewhere. Being in debt due to escalation is also a prominent feature that was identified by almost all models to be a risk factor. For this outcome, going from no debt to being in debt due to escalation, the odds of eating elsewhere increases multiplicatively by 3. Additionally, the odds of eating elsewhere among households whose priority needs is food is 2.13 times those who eat elsewhere but their top needs are not food. These households are probably more in need of food to satisfy their basic nutritional needs, which may force them to eat somewhere else because there isn't sufficient food at home or they don't have enough cash to buy it [13].

The final bar graph 3.16 for *asking strangers* illustrates an expected association between debt due to escalation, food expenditures, challenges in obtaining money and cost of medication barrier with the outcome asking strangers. These four features are present in six out of eight features indicating that most models found that they significantly affect the outcome. Household characteristics such as number of single members, number of seniors in household and number of visually impaired in household along side proximity to frontline features make a strong appearance

in the graph of best features. Moreover, lack of work and opportunity and having short-term expenditures are possible preditors that lead household to cope by asking stranger for help. Table C.7 also illustrates the importance of the predictors debt due to escalation and challenges obtaining money when it comes to asking stranger. A very notable and severe association is going from no debt to being in debt due to escalation, the odds of asking strangers increases multiplicatively by 7. We expect these results, as people often turn to one another for mental, financial, and social support during times of war. Also, households going from relative to severe food insecurity, the odds of asking stranger increases multiplicatively by 4.38. As for the household characteristics, the positive coefficients of the features number of single household members and count of visually disabled people are risk factors. In a household, with every one member (single or disabled) increase, the odds of asking strangers increases multiplicatively by 1.24 and 1.32 respectively. Furthermore, the closer the household to Volnovaskyi, located in Donetsk Oblast, the more the odds of asking strangers increase. Going from no presence of explosive ordnance to presence of explosive ordnance in communities, the odd of asking strangers increases by 1.78. Finally, the odds of asking strangers among those who do not want to receive humanitarian assistance, is 0.24 times the odds of asking strangers among those who want to receive humanitarian assistance.

Chapter 4

Discussion and Conclusion

This research consists of two main objectives: to find the optimal set of features that best explain the outcome while handling the data imbalance and to quantify an association between the predictors and the negative coping mechanisms. In our evaluation of the predictive models, we employed a downsampling technique across all models and utilized a cost-sensitive learning approach specifically for the random forest model. In the analysis provided in the results section for model performance, we notice RF outperforms all models, before and after tuning, when handling the following outcomes: *reduce health expenditures*, *spent savings* and *borrow food*. As we reach the *degrading income source* and *reduce education expenditures* outcomes, we notice that the XGB and TabNet models emerge as a competitor for the RF model. Then at the level of *sell household assets*, *eat elsewhere* and *ask stranger* outcomes, TabNet consistently performs best, scoring the highest *AUC* and *accuracy* scores. In summary, RF demonstrated the most consistent performance among all the models, XGB served as a competitor to RF, TabNet performed well on specific outcomes only and Lasso lagged other models in performance but still delivered solid results overall. We suspect that the quality of the data is the primary factor dictating model performance, especially considering that downsampling leads to the loss of valuable instances from the dataset. A key limitation is that TabNet may not have achieved its best performance across all models due to the manual tuning process. For instance, take the highlighted result in table 3.4 where the AUC score before tuning (0.74) was higher than that of the tuned model (0.71). Its computational complexity and the wide range of parameters available for tuning makes it challenging to obtain optimal settings. Consequently, some TabNet baseline models performed slightly better than the tuned ones. With proper tuning, we believe TabNet might have shown outstanding results.

Throughout this research, we found valuable insights that validated and added onto the existing literature. In the literature, several factors were identified as predictors of negative coping mechanisms. They were mainly concentrated around the following themes: household characteristics, proximity to frontline and economic conditions of household during times of conflict. Machine learning techniques proved to achieve results that align closely with those reported in the literature, which were derived through qualitative or observational methods. Building upon our research

findings, we have taken the initiative to further explore and quantify the identified predictors. The two most common predictors of all the negative coping mechanisms were debt due to escalation and challenges obtaining money. They're both risk factors related to the financial status of a household during war. In accordance to our results for *reducing health expenditures*, an observational study by Ekzayez et al.,[18] tackling the impact of armed conflict on utilisation of health services in Syria, found an association between armed conflicts and utilization of health care services. They showed that financially struggling households that live in affected areas limit their access to health services. So, as the household struggles to secure money the more it is likely to reduce health care expenses which also happens to align with Namasivayam et al.'s [31] insights. These households will find it unsafe or impossible to travel to healthcare facilities aligning with Ekzayez et al.'s claim. One could elaborate on this matter looking into the destruction of infrastructure where shelling can damage healthcare facilities consequently reducing the availability of healthcare services due to security concerns [29]. On the other hand, we found that households experiencing no security concerns, no utility service interruption, no living condition issues, no assistance barriers and no conflict damage are less likely to adopt harmful coping behaviors such as *borrowing food* or *reducing education expenditures*. In the context of healthcare, our models were able to show that households struggling to afford medicine or to access healthcare services due to their cost, are at a higher chance of coping which comes in accordance to the findings of 'REACH Health Sector Needs Assessment' in Ukraine [32]. Based on the findings reported by Justino (2011) [23], Hill et al. (2019) [22], and Tongruksawattana et al. (2010) [26], who discussed the vulnerability of the household from a financial perspective, the higher the income of a household, the less likely they'll spend their savings. Our findings suggested that not only is a higher income a protective factor of *spending savings*, but also from *borrowing food*, *degrading income source*, *reducing education expenses* and *eating elsewhere*. During conflict households often become trapped in a cycle of debt and sometimes rely on multiple sources of income [12], in turn these household are in need for financial assistance to cover the debt payments. In our research we found that these households tend to *borrow food* and *ask strangers* for help and support more. Another significant indicator in this study was the reduced coping strategy index explored by Maxwell and Caldwell (2008) [13] where we mentioned an association between rCSI and *borrowing food*/*eating elsewhere*. We build on existing knowledge the association between coping and food insecurity where we found rCSI was associated positively (risk factor) with all the target variables in this study except *spending savings*. In regards to the proximity to frontline features, almost all significant features showed a distinctive association where the farthest the household is from the frontline, the more likely they'll cope. This observation could be due to other regions of Ukraine being targeted, not limited to the frontline areas such as the capital Kyiv. The insights on gender distribution within the household were limited where only two models identified associations with the gender. While research found that females are more likely to cope by *asking relatives* for help [24], we found that as the number of females increases by one, the household becomes more likely to *borrow food*. Additionally, with each additional male in the household,

the likelihood of *eating elsewhere* increases. Other reports suggest that women cope by contributing from their *savings* or *selling their gold* pieces [12], however, we did not achieve such association. Nonetheless, other household characteristics served as predictors of harmful coping behaviors adopted by households during conflict. Results show that the increased number of children and seniors was accompanied by lower likelihood of coping. Specifically, the number of seniors in a household is a protective factor of *spending savings*, *degrading income source*, *selling household assets*, *eating elsewhere* and *reducing education expenses*. Moreover, we suggest that the bigger the household the more likely it is to *reduce education expenditures* and the higher the number of single members the more likely it is to seek help from others or degrade their income source. Finally, we highlight the role of visual and mobility disabilities as significant risk factors for negative coping strategies.

With nearly half of the Ukrainian population in need, these findings emphasize the importance of understanding the specific vulnerabilities and strengths present within diverse household compositions when designing targeted interventions. From a humanitarian perspective, recognizing the challenges faced by different household allows for the creation targeted interventions that can directly address the needs of the most vulnerable households in conflict settings. During conflict, resources are often limited and the consequences of negative coping mechanisms can be severe on the long run. Targeted interventions that take into consideration the identified household characteristics in this research can help prevent these negative outcomes, consequently contributing to the development of stable, less vulnerable community.

Moreover, the integration of machine learning techniques in this research provides a powerful tool for enhancing our understanding of household vulnerability. By analyzing the complex, non-linear and big data that we put together and identifying predictors of negative coping mechanisms that might not be evident through traditional or observational methods, machine learning enables more accurate predictions. This in turn accounts for the development of more effective intervention programs. In conclusion, this research highlights the necessity of targeted interventions that are informed by a deep understanding of the household compositions and vulnerabilities through machine learning. The application of machine learning not only enriches the analysis but also paves the way for more informed decision-making in humanitarian efforts. This research adds to the expanding knowledge that attempts to enhance the lives of people in need and offers insights that can inform policy and practice.

Appendix A

GIS Data

GIS Data Description

Two datasets are used: one defines the oblast boundaries and the other defines the raion boundaries. Data can be found in: Source-1 and Source-2. This data is stored as a vector that defines the geometry of Ukraine's territory in the format of geometry vectors consisting of polygons that outline the boundaries of *Oblasts* and *Raions*.

Oblast Administrative Boundaries

	shapeName	shapeISO	shapeID	shapeGroup	shapeType	geometry
0	Kherson Oblast	UA-65	14850775B89584667307707	UKR	ADM1	POLYGON ((35.23342 45.79173, 35.23312 45.79380...
1	Volyn Oblast	UA-07	14850775B62859603218954	UKR	ADM1	POLYGON ((25.11276 50.28727, 25.11147 50.29428...
2	Rivne Oblast	UA-56	14850775B55533755566407	UKR	ADM1	POLYGON ((25.11276 50.28727, 25.11291 50.28489...
3	Zhytomyr Oblast	UA-18	14850775B32159522680847	UKR	ADM1	POLYGON ((27.19595 50.56224, 27.19597 50.56181...
4	Kyiv Oblast	UA-32	14850775B9435746000194	UKR	ADM1	MULTIPOLYGON (((30.34907 50.48887, 30.34605 50...
5	Chernihiv Oblast	UA-74	14850775B4016712610052	UKR	ADM1	POLYGON ((32.14266 50.34881, 32.16318 50.35177...

Figure A.1: Snapshot of the Oblast Administrative Boundaries Dataset

The above dataset includes:

- **shapeName** contains the English name of each oblast.
- **shapeISO** contains the ISO code for each oblast. It consists of the code "UA" followed by a unique number for each region (oblast).
- **shapeID** contains a unique ID for each oblast.
- **shapeGroup** indicates that we are within the boundaries of Ukraine by displaying the value "UKR".
- **shapeType** indicates the level of the boundaries in our dataset. "ADM1" indicates that we are at the oblast level.
- **geometry** contains the polygon and multi-polygon vectors, mentioned above, that define the boundaries of the oblasts.

Raion Administrative Boundaries

	shapeName	shapeISO	shapeID	shapeGroup	shapeType	geometry
0	Вознесенський район		31201334B26086675319741	UKR	ADM2	POLYGON ((30.77689 47.64738, 30.78831 47.64904...
1	Лубенський район		31201334B69155138870911	UKR	ADM2	POLYGON ((32.68880 49.66026, 32.69220 49.66209...
2	Берегівський район		31201334B93378394099264	UKR	ADM2	POLYGON ((23.26701 48.10030, 23.26754 48.10101...
3	Дніпровський район		31201334B39514660848430	UKR	ADM2	POLYGON ((34.18541 47.97009, 34.18843 47.96100...
4	Старобільський район		31201334B95322055593201	UKR	ADM2	POLYGON ((38.81781 49.05478, 38.81745 49.05935...

Figure A.2: Snapshot of the Raion Administrative Boundaries Dataset

The above dataset includes:

- **shapeName** contains the Ukrainian name of each raion.
- **shapeISO** column is missing the ISO codes.
- **shapeID** contains a unique ID for each raion.
- **shapeGroup** indicates that we are within the boundaries of Ukraine by displaying the value "UKR".
- **shapeType** indicates the level of the boundaries in our dataset. "ADM2" indicates that we are at the raion level.
- **geometry** contains the polygon and multi-polygon vectors, mentioned above, that define the boundaries of the raions.

Challenges with the Raion Boundaries Dataset

While working on the Raion dataset, we encountered several challenges. The most apparent one was the missing ISO codes of all the Raions and the language of the **shapeName** column which is in Ukrainian. ISO codes are very important since in the MSNA dataset Oblasts and Raions are referred to by their respective codes. To remedy this, the English Raion names, Ukrainian Raion names and their respective ISO codes included in the "KOBO Choices" sheet were extracted into a separate dataset as shown in figure A.3.

	A	B	C
1	name	English	Ukrainian
2	UA4802	Bashtanskyi	Баштанський
3	UA1802	Berdychivskyi	Бердичівський
4	UA2102	Berehivskyi	Берегівський
5	UA5102	Berezivskyi	Березівський
6	UA6502	Beryslavskyi	Бериславський
7	UA5104	Bilhorod-Dnistrovsky	Білгород-Дністровський
8	UA3202	Bilotserkivskyi	Білоцерківський
9	UA6302	Bohodukhivskyi	Богодухівський
10	UA5106	Bolhradskyi	Болградський

Figure A.3: Snapshot of the Raion Names & ISO codes Dataset

Next step was to merge the "Raion Names & ISO codes Dataset" with "Raion Administrative Boundaries Dataset" on the Ukrainian name of each Raion since it's a common column between both datasets. hence, we get the following dataset:

	shapeName	shapeISO	shapeID	shapeGroup	shapeType	geometry	name	English
0	Вознесенський		31201334B26086675319741	UKR	ADM2	POLYGON ((30.77689181704463 47.647378131383576...	UA4804	Voznesenskyi
1	Лубенський		31201334B69155138870911	UKR	ADM2	POLYGON ((32.68880148446695 49.66026109754716,...	UA5304	Lubenskyi
2	Берегівський		31201334B93378394099264	UKR	ADM2	POLYGON ((23.267008860270767 48.10030294042841...	UA2102	Berehivskyi
3	Дніпровський		31201334B39514660848430	UKR	ADM2	POLYGON ((34.18540518625563 47.970093116562246...	UA1202	Dniprovskyi
4	Старобільський		31201334B95322055593201	UKR	ADM2	POLYGON ((38.8178139480211 49.054781002240446,...	NaN	NaN

Figure A.4: Snapshot of the Merged Dataset

It is important to note that the columns **name** and **English**, which display the ISO code and English name respectively, have some missing values "NaN". This is due to fact that the "Raion Names & ISO codes Dataset" only has information on Raions that were included in the study while "Raion Administrative Boundaries Dataset" holds information on all the Raions. Then, the "NaN" in the **English** column were imputed manually with their English name for easier identification and representation of the Raions as shown in the interactive map in figure A.5.

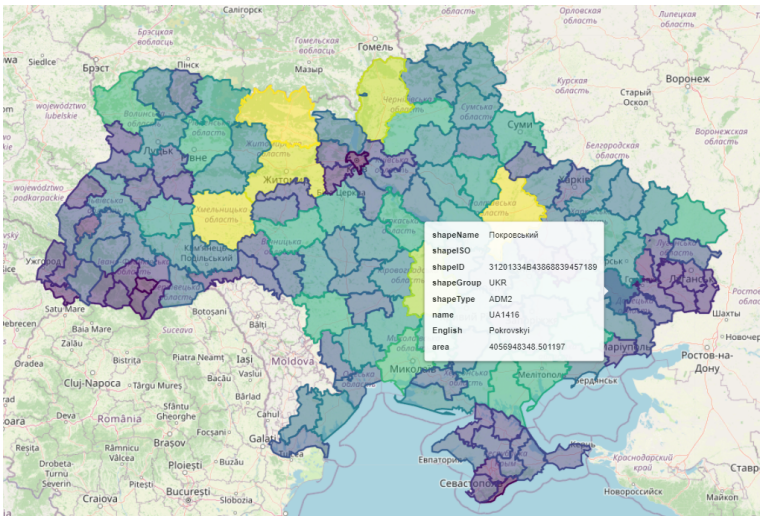


Figure A.5: Interactive map of Ukraine at the Raion Level

Another problem surfaced when visualizing the raions. We noticed a missing geographical boundary representation of 'Kyiv'. To address this issue, we extracted data from *OpenStreetMaps* projects [57]. This site has GIS (Geographic Information System) data for mapping. It offers geospatial data for each Oblast or Raion separately. So, we extracted the data found on 'Kyiv' and added it to the dataset Source-3.

Figure A.6: Adding Kyiv to the Ukrainian map

Distances from Frontline

As of December 2023, the identified front-lines, on the southern east side of Ukraine, include following oblasts: Luhansk, Donetsk, Zaporizhzhia and Kherson. The identified front-line is dependent on the development of conflict on the ground. Using geo-spatial data and the identified front-line, the distance between a point on the front-line and an Oblast or Raion, is calculated. First, the centroid of each geometry is calculated at the Oblast and raion level A.7 [58].

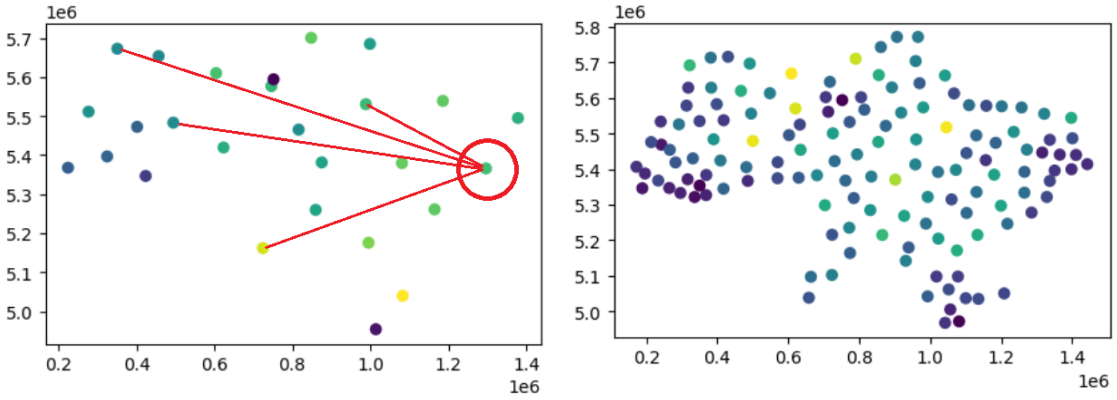


Figure A.7: Map representation of Ukrainian Oblasts (on the left) and Raions (on the right) using their corresponding centroids

Then, after we calculated the distance of each oblast or raion from the front-line, the ISO codes in the datasets are replaced with the distances. These ISO codes are indicators of the households' location.

Appendix B

Model Parameters

Model Parameter Selection

Spent Savings

Model	Parameter	5 Fold CV	10 Fold CV
Lasso	C	0.01	0.01
	Solver	'saga'	'saga'
XGB	N_estimators	100	200
	Max Depth	7	7
	Learning rate	0.1	0.1
	Subsample	0.9	0.9
	Gamma	0.01	1
	Alpha	10	10
Random Forest	N_estimators	200	200
	Max Depth	None	20
	Min Sample Split	2	2
	Min Sample Leaf	2	1
	Max features	'sqrt'	'sqrt'
TabNet	n_d	40	40
	n_a	40	40
	n_steps	3	1
	Momentum	0.4	0.06
	Clip	2	2
	Learning Rate	default	0.1

Table B.1: Parameter settings for different models across cross-validation folds for Spend Savings

Borrowed Food

Model	Parameter	5 Fold CV	10 Fold CV
Lasso	C	0.1	0.1
	Solver	'saga'	'saga'
XGB	N_estimators	100	200
	Max Depth	3	2
	Learning rate	0.1	0.1
	Subsample	0.8	0.9
	Gamma	0.01	0.1
	Alpha	1	1
Random Forest	N_estimators	200	100
	Max Depth	None	None
	Min Sample Split	2	2
	Min Sample Leaf	2	2
	Max features	'sqrt'	'sqrt'
TabNet	n_d	36	40
	n_a	36	40
	n_steps	1	1
	Momentum	0.01	0.01
	Learning Rate	0.1	0.1

Table B.2: Parameter settings for different models across cross-validation folds for Borrowing Food

Degrade Income Source

Model	Parameter	5 Fold CV	10 Fold CV
Lasso	C	0.1	0.1
	Solver	'liblinear'	'liblinear'
XGB	N_estimators	50	50
	Max Depth	7	7
	Learning rate	0.1	0.1
	Subsample	0.8	0.8
	Gamma	0.01	0.01
	Alpha	0.1	0.1
Random Forest	N_estimators	100	100
	Max Depth	10	None
	Min Sample Split	2	5
	Min Sample Leaf	1	2
	Max features	'sqrt'	'sqrt'
TabNet	n_d	24	36
	n_a	24	36
	n_steps	1	1
	Momentum	0.05	0.01
	Clip	2	2

Table B.3: Parameter settings for different models across cross-validation folds for Degrading Income Source

Reduce Education Expenditures

Model	Parameter	5 Fold CV	10 Fold CV
Lasso	C	0.01	0.01
	Solver	'saga'	'saga'
XGB	N_estimators	100	50
	Max Depth	2	2
	Learning rate	0.1	0.2
	Subsample	0.9	0.8
	Gamma	0.01	1
	Alpha	0.01	0.1
Random Forest	N_estimators	100	20
	Max Depth	10	None
	Min Sample Split	2	5
	Min Sample Leaf	1	1
	Max features	'sqrt'	'sqrt'
TabNet	n_d	30	24
	n_a	30	24
	n_steps	1	1
	Momentum	0.04	0.4
	Clip	2	default

Table B.4: Parameter settings for different models across cross-validation folds for Reducing Education Expenditures

Sell HH Assets

Model	Parameter	5 Fold CV	10 Fold CV
Lasso	C	0.1	0.1
	Solver	'saga'	'saga'
XGB	N_estimators	50	20
	Max Depth	3	5
	Learning rate	0.1	0.2
	Subsample	0.9	0.8
	Gamma	1	1
	Alpha	1	1
Random Forest	N_estimators	200	100
	Max Depth	10	10
	Min Sample Split	5	2
	Min Sample Leaf	2	1
	Max features	'sqrt'	'sqrt'
TabNet	n_d	24	30
	n_a	24	30
	n_steps	1	1
	Momentum	0.3	0.04
	Learning Rate	0.1	0.1

Table B.5: Parameter settings for different models across cross-validation folds for Selling HH Assets

Eat Elsewhere

Model	Parameter	5 Fold CV	10 Fold CV
Lasso	C	0.1	0.01
	Solver	'saga'	'liblinear'
XGB	N_estimators	50	50
	Max Depth	5	5
	Learning rate	0.1	0.1
	Subsample	0.8	0.8
	Gamma	0.01	1
	Alpha	0.01	1
Random Forest	N_estimators	200	100
	Max Depth	10	None
	Min Sample Split	2	2
	Min Sample Leaf	1	1
	Max features	'sqrt'	'log2'
TabNet	n_d	24	30
	n_a	24	30
	n_steps	1	1
	Momentum	0.4	0.4
	Learning Rate	default	0.01

Table B.6: Parameter settings for different models across cross-validation folds for Eating Elsewhere

Ask Stranger

Model	Parameter	5 Fold CV	10 Fold CV
Lasso	C	0.1	0.1
	Solver	'saga'	'saga'
XGB	N_estimators	100	100
	Max Depth	2	2
	Learning rate	0.1	0.1
	Subsample	0.8	0.9
	Gamma	0.1	0.1
	Alpha	1	1
Random Forest	N_estimators	200	100
	Max Depth	None	None
	Min Sample Split	2	5
	Min Sample Leaf	2	2
	Max features	'sqrt'	'sqrt'
TabNet	n_d	30	30
	n_a	30	30
	n_steps	1	1
	Momentum	0.3	0.04
	Learning Rate	0.1	0.1

Table B.7: Parameter settings for different models across cross-validation folds for Asking Strangers

Appendix C

Lasso Model Summary & Learning Curves

LASSO Model Summary

	Coefficient	Std. Error	Z-Score	P-Value
Intercept	-0.104700	0.024800	-4.222800	0.000000
Long-term Expenditures for Domestic Consumption	0.300400	0.009500	31.468600	0.000000
Healthcare Desired: Chronic, Non-communicable illness	0.043200	0.012000	3.608900	0.000300
Income	-0.043300	0.011700	-3.703200	0.000200
Intention to Access Healthcare Service	0.231900	0.012100	19.212300	0.000000
Medicine Barrier: Cost of Medication	0.051300	0.010000	5.154500	0.000000
Food Expenditures	0.031000	0.011400	2.728400	0.006400
Number of Seniors in HH	-0.104100	0.010300	-10.062800	0.000000
Number of Employed in HH	-0.099700	0.011300	-8.831000	0.000000
HH distance from Kalmiusky	0.178400	0.009500	18.870000	0.000000
Challenges Obtaining Money	0.372600	0.028000	13.305100	0.000000
Challenges Obtaining Money: Price Increase, Static Income	0.258600	0.028500	9.076400	0.000000
Decline Assistance	-0.443400	0.023500	-18.832300	0.000000
Debt Due to Escalation	0.341900	0.026600	12.874100	0.000000
Security Concerns: None	-0.303400	0.023800	-12.758200	0.000000

Table C.1: Spent Savings Lasso Summary Table

	Coefficient	Std. Error	Z-Score	P-Value
Intercept	-1.177700	0.052900	-22.279700	0.000000
Medicine Barrier: Cost of Medication	0.139900	0.010200	13.680300	0.000000
Long-term Expenditures for Domestic Consumption	0.195800	0.009500	20.596900	0.000000
Income	-0.201600	0.010700	-18.920000	0.000000
Intention to Access Healthcare Service	0.192900	0.026800	7.203500	0.000000
Sought Healthcare Services	-0.145300	0.026400	-5.511800	0.000000
HH Decision Making: Money	-0.074500	0.011600	-6.419700	0.000000
HH distance from Donetsk	-0.171700	0.009700	-17.740900	0.000000
Sought fever Medicine	0.083900	0.010100	8.288500	0.000000
Number of Females in HH	0.033600	0.011400	2.955400	0.003100
Reduced Coping Strategy Index 1	0.693900	0.022500	30.806800	0.000000
Reduced Coping Strategy Index 2	1.441400	0.056200	25.659700	0.000000
Debt Due to Escalation	1.260600	0.027900	45.247900	0.000000
Priority Needs: Financial Assistance for Debt	0.816500	0.042300	19.312500	0.000000
Challenges Obtaining Money	0.816500	0.032600	25.024900	0.000000
Missing Non-Food Items: None	-0.287900	0.022900	-12.557400	0.000000
Living Condition Issues: None	-0.436100	0.035800	-12.193100	0.000000
Hygiene Item Availability: Everything was affordable	-0.165200	0.025700	-6.430900	0.000000
Missing Non-Food Items: Winter Clothes	0.222700	0.037800	5.895200	0.000000
Declined Assistance	-0.607300	0.024200	-25.132500	0.000000
Garbage Disposal: Burying	0.417300	0.033300	12.519000	0.000000
Priority Needs: Livelihood Support Employment	0.460300	0.028300	16.261900	0.000000
Children Services: Supportive Group Activities	0.204000	0.023600	8.657000	0.000000
Challenges Obtaining Money: Lack of Work Opportunity	0.220800	0.035200	6.267500	0.000000
Security Concerns: None	-0.132100	0.024100	-5.491600	0.000000
Sanitation Facility: Flush to Pit Latrine	0.200600	0.021200	9.455300	0.000000
Challenges Obtaining Money: Low Pension Benefits	-0.198300	0.030000	-6.604500	0.000000
Garbage Disposal: Collection Point - Regularly	0.090300	0.028700	3.149500	0.001600
Challenges Obtaining Money: Low Salaries	0.249000	0.030200	8.234800	0.000000

Table C.2: Borrowed Food Lasso Summary

	Coefficient	Std. Error	Z-Score	P-Value
Intercept	-0.112700	0.058700	-1.920000	0.054900
Long-term Expenditures for Domestic Consumption	0.135800	0.009600	14.170500	0.000000
Number of Single in HH	0.208900	0.015000	13.964700	0.000000
Food Expenditures	0.166700	0.011600	14.401900	0.000000
Income	-0.151300	0.011500	-13.204600	0.000000
Number of Seniors in HH	-0.402700	0.011000	-36.717900	0.000000
Number of Children Enrolled	-0.039500	0.018400	-2.142100	0.032200
Sought Healthcare Service	0.040100	0.010600	3.793900	0.000100
Number of Children in HH	-0.174300	0.021300	-8.173700	0.000000
Number of Disability Mobility in HH	0.109100	0.010000	10.938000	0.000000
Sought Fever Medicine	-0.031400	0.010100	-3.109500	0.001900
Medicine Barrier: Cost of Medication	0.044700	0.010200	4.376800	0.000000
HH distance from Henicheski	0.087100	0.009600	9.065900	0.000000
Income Sources: Informal Employment	1.539900	0.030100	51.111700	0.000000
Debt Due to Escalation	0.868200	0.027100	31.979900	0.000000
Challenges Obtaining Money: Lack of Work Opportunity	0.672600	0.034700	19.406600	0.000000
No Legal Assistance Requirement	-0.702500	0.028300	-24.837000	0.000000
Challenges Obtaining Money	0.592800	0.022000	26.998900	0.000000
No Conflict Damage	-0.483900	0.035400	-13.650900	0.000000
Explosives Affecting Livelihoods	0.755800	0.036800	20.562400	0.000000
Utility Services: No interruption	-0.737400	0.025400	-28.981400	0.000000
Reduced Coping Strategy Index 1	0.522300	0.022100	23.654400	0.000000
Reduced Coping Strategy Index 2	1.174000	0.056700	20.705500	0.000000
Main Food Source: In-kind Food Aid	0.775100	0.037600	20.627600	0.000000
Living Conditions Issues: None	-0.632400	0.035800	-17.669400	0.000000
Income Source: Irregular Employment	0.885100	0.036200	24.418100	0.000000
Shelter Type: Detached House	0.134800	0.021700	6.224200	0.000000
No Children Services	0.213800	0.024400	8.769600	0.000000
Priority Needs: Hygiene Non-Food Items	-0.107100	0.027300	-3.915400	0.000100
Income Source: Regular Employment	-0.052200	0.022300	-2.344700	0.019000
Priority Needs: Livelihoods Support Employment	0.554200	0.028400	19.523400	0.000000

Table C.3: Degrade Income Source Lasso Summary

	Coefficient	Std. Error	Z-Score	P-Value
Intercept	0.000000	0.043000	0.000000	1.000000
Health Service Access	0.362500	0.017200	21.083600	0.000000
Medicine Barrier: cost of medication	0.218600	0.010400	21.011500	0.000000
HH size	0.210500	0.014000	15.017800	0.000000
Income	-0.131800	0.010500	-12.591100	0.000000
Long-term Expenditures for Domestic Consumption	0.258100	0.009500	27.211400	0.000000
HH distance from Beryslavskyi	0.201100	0.009500	21.192600	0.000000
Number of Seniors in HH	-0.134700	0.009800	-13.765800	0.000000
No Barriers Accessing Healthcare	-0.238600	0.016300	-14.667200	0.000000
Number of Children Enrolled	0.372400	0.012900	28.969500	0.000000
Challenges Obtaining Money	0.834600	0.021900	38.065700	0.000000
Reduced Coping Strategy Index 1	0.638600	0.022300	28.648700	0.000000
Reduced Coping Strategy Index 2	1.102300	0.055800	19.764100	0.000000
No Legal Assistance Required	-0.467000	0.029200	-16.007300	0.000000
Challenges Obtaining Money: Lack of Work Opportunity	0.191000	0.034200	5.580900	0.000000
Hygiene Item Availability: Everything was affordable	-0.079100	0.024100	-3.280900	0.001000
No Security Concerns	-0.643100	0.023800	-27.027000	0.000000
No Assistance Barriers	-0.474900	0.019100	-24.859600	0.000000
Water Safety	-0.530400	0.021200	-25.074600	0.000000
Priority Needs: Livelihood Support Employment	0.548300	0.028300	19.383600	0.000000
Government Social Services	0.230400	0.032900	7.009500	0.000000

Table C.4: Reduce Education Expenditures Lasso Summary

	Coefficient	Std. Error	Z-Score	P-Value
Intercept	-0.572000	0.060000	-9.540400	0.000000
Medicine Barrier: Cost of Medication	0.103700	0.010900	9.501000	0.000000
Long-term Expenditures for Domestic Consumption	0.053400	0.009500	5.592000	0.000000
Number of Seniors in HH	-0.149600	0.010700	-13.933300	0.000000
HH distance from Khersonskyi	0.105700	0.012300	8.577300	0.000000
HH distance from Svativskyi	0.271300	0.012100	22.468500	0.000000
Intention to Access Healthcare Service	0.117800	0.010200	11.598400	0.000000
Food Expenditures	0.038800	0.011600	3.352000	0.000800
Short-term Expenditures for Domestic Consumption	0.046700	0.011200	4.154800	0.000000
Barrier Prevented Healthcare Access: Cost	0.087700	0.010400	8.397700	0.000000
Number of Employed in HH	0.083200	0.013600	6.135800	0.000000
Debt Due to Escalation	1.389500	0.027300	50.840200	0.000000
Hygiene Item Availability: Everything was affordable	-0.167300	0.025000	-6.697700	0.000000
Challenges Obtaining Money	0.766500	0.025800	29.762400	0.000000
No Legal Assistance Required	-0.549300	0.029600	-18.568100	0.000000
Challenges Obtaining Money: Lack of Work Opportunity	0.412200	0.033600	12.278200	0.000000
No Interruptions in Utility Services	-0.842000	0.038400	-21.933700	0.000000
Declined Assistance	-0.990900	0.025300	-39.208600	0.000000
Internet Coverage 1	0.706900	0.041000	17.256900	0.000000
Internet Coverage 2	0.230800	0.035400	6.517000	0.000000
Missing Non-Food Items: Blanket	-0.094600	0.040800	-2.317500	0.020500
Income Source: Regular Employment with Salary	-0.734900	0.026000	-28.241100	0.000000
Reduced Coping Strategy Index 1	0.433100	0.022500	19.275900	0.000000
Reduced Coping Strategy Index 2	0.826400	0.056700	14.578400	0.000000
Main Source Drinking: Bottled Water	0.256600	0.027900	9.198500	0.000000
Challenges Obtaining Money: Low Salaries	0.344300	0.030900	11.137100	0.000000
Preferred Assistance: Services	0.165700	0.022400	7.383900	0.000000
Agricultural Land Purpose: Production for Consumption	-0.310000	0.019500	-15.900700	0.000000
Assistance Barriers: No Info on Assistance Registration	0.114800	0.024500	4.679200	0.000000
Internet Access: Mobile	0.274200	0.027900	9.817400	0.000000
Priority Needs: Repairs in Damaged Accommodation	0.455000	0.035100	12.975100	0.000000

Continued on next page

	Coefficient	Std. Error	Z-Score	P-Value
Government Social Services	0.230400	0.032900	7.013800	0.000000
Utility Service Interruption: Main Electricity	-0.145000	0.046300	-3.131600	0.001700

Table C.5: Sell HH Assets Lasso Summary

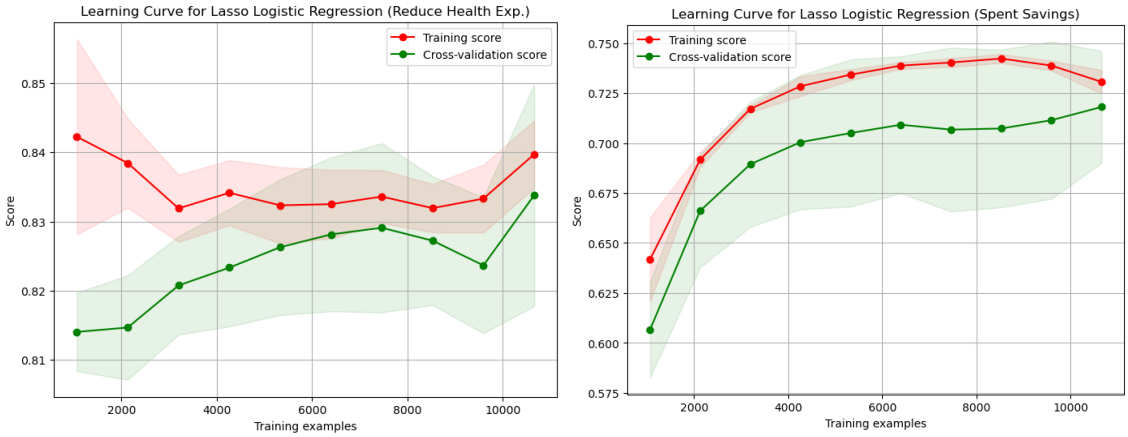
	Coefficient	Std. Error	Z-Score	P-Value
Intercept	0.000000	0.125900	0.000000	1.000000
Income	-0.481700	0.012000	-40.296300	0.000000
Decision Making: Migration	-0.594400	0.011900	-49.775300	0.000000
Food Expenditures	-0.209900	0.011800	-17.713700	0.000000
HH distance from Melitopolskyi	0.277400	0.009800	28.169400	0.000000
Short-term Expenditures for Domestic Consumption	0.068200	0.011600	5.885200	0.000000
Barrier Prevented Healthcare Access: Cost	0.132200	0.009500	13.924000	0.000000
Sought Blood Pressure Medicine	-0.083100	0.010600	-7.864300	0.000000
Number of Mobility Disability in HH	0.264400	0.009900	26.828900	0.000000
Number of Seniors in HH	-0.307100	0.010900	-28.074900	0.000000
Number of Males in HH	0.235100	0.011800	19.871300	0.000000
Reduced Coping Strategy Index 1	0.830200	0.022400	37.088400	0.000000
Reduced Coping Strategy Index 2	1.864200	0.056300	33.107500	0.000000
Debt Due to Escalation	1.095000	0.027100	40.466300	0.000000
Challenges Obtaining Money	0.366500	0.021200	17.254300	0.000000
Priority Needs: Food	0.757300	0.020400	37.069100	0.000000
Hygiene Item Availability: Everything was affordable	-0.315400	0.024300	-12.983500	0.000000
No legal Assistance Required	-0.894300	0.028300	-31.602100	0.000000
Main Heating Source: Wood	0.176700	0.021600	8.170900	0.000000
No Property Concerns	-0.758800	0.040800	-18.610900	0.000000
Preferred Assistance: Cash	0.335100	0.020900	16.050100	0.000000
Shelter Type: Detached House	-0.137500	0.025800	-5.336700	0.000000
Main Source Drinking Water: Personal Borehole or Well	-0.140000	0.022000	-6.372700	0.000000
Water Sufficiency: Cooking	-0.331100	0.064700	-5.114600	0.000000
No Children Concerns	-0.284000	0.021600	-13.152500	0.000000
Domestic Space Issues: Food & Water Storage	-0.393300	0.104500	-3.763300	0.000200
Receive Food Assistance	0.806600	0.027400	29.384500	0.000000
Main Source Drinking: Bottled Water	-0.628900	0.028200	-22.304500	0.000000

Table C.6: Eat Elsewhere Lasso Summary

	Coefficient	Std. Error	Z-Score	P-Value
Intercept	-1.438000	0.075300	-19.099100	0.000000
Food Expenditures	-0.475600	0.011800	-40.216000	0.000000
Medicine Barrier: Cost of Medication	0.195300	0.010500	18.571100	0.000000
Short-term Expenditures for Domestic Consumption	-0.159200	0.011700	-13.646900	0.000000
No Medicine Barrier	-0.109200	0.010400	-10.527300	0.000000
Number of Single in HH	0.213400	0.010900	19.558300	0.000000
Number of Seniors in HH	-0.406300	0.010900	-37.232200	0.000000
Income	0.085300	0.012300	6.957900	0.000000
Decision Making: Migration	-0.147000	0.011100	-13.298600	0.000000
HH Distance from Volnovaskyi	0.188100	0.009600	19.643900	0.000000
Number of Visual Disability in HH	0.279000	0.009500	29.281600	0.000000
Debt Due to Escalation	1.964000	0.028200	69.735800	0.000000
Challenges Obtaining Money	0.595600	0.022200	26.781300	0.000000
Reduced Coping Strategy Index 1	0.502200	0.022600	22.233900	0.000000
Reduced Coping Strategy Index 2	1.479100	0.056300	26.269000	0.000000
Hygiene Item Availability: Washing Soap	0.522100	0.039700	13.142200	0.000000
Challenges Obtaining Money: Lack of Work Opportunity	0.150500	0.033500	4.498200	0.000000
No Legal Assistance Required	-0.788800	0.028100	-28.066300	0.000000
Water Sufficiency: Cooking	-0.637800	0.064300	-9.918100	0.000000
Sanitation Facility: Pit Latrine with Slab	0.459700	0.020300	22.596400	0.000000
Explosives Affecting Livelihoods	0.574600	0.035400	16.210200	0.000000
Aid Providers Information: None	-0.558100	0.022600	-24.707700	0.000000
Internet Access: Do Not Use	0.450900	0.028000	16.126800	0.000000
Hygiene Item Availability: Everything was affordable	0.126200	0.033000	3.829100	0.000100
Declined Assistance	-1.414900	0.026000	-54.452400	0.000000
Aid Providers Information: How to Access Food Assistance	0.369600	0.028600	12.909200	0.000000
Priority Needs: Financial Assistance to repay Debt	0.161600	0.042600	3.792800	0.000100

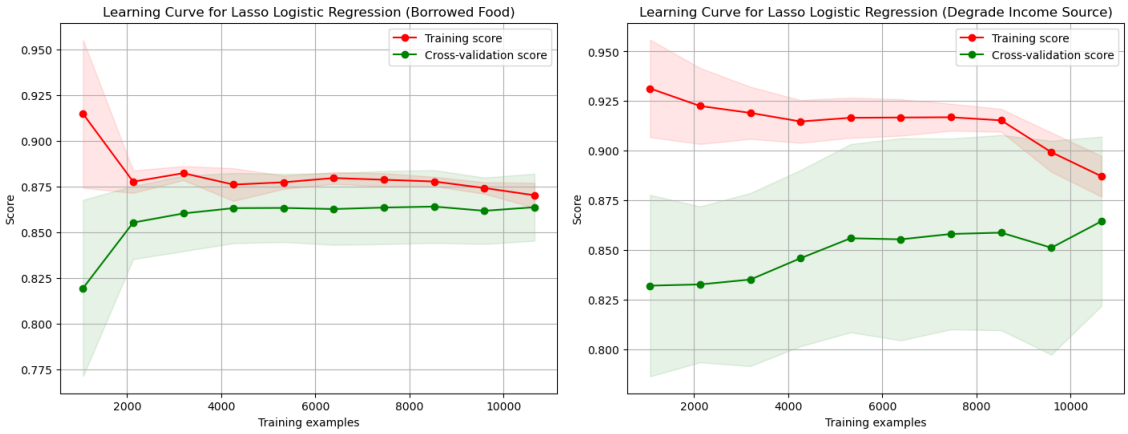
Table C.7: Ask Stranger Lasso Summary

Learning Curves of LASSO Model



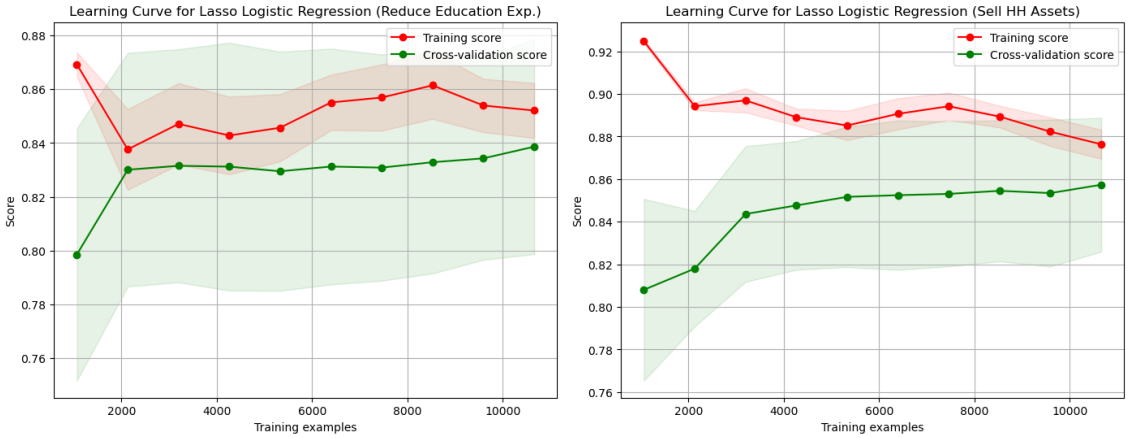
(a) (b)

Figure C.1: Learning Curves for Reduce Health Exp. and Spent Savings



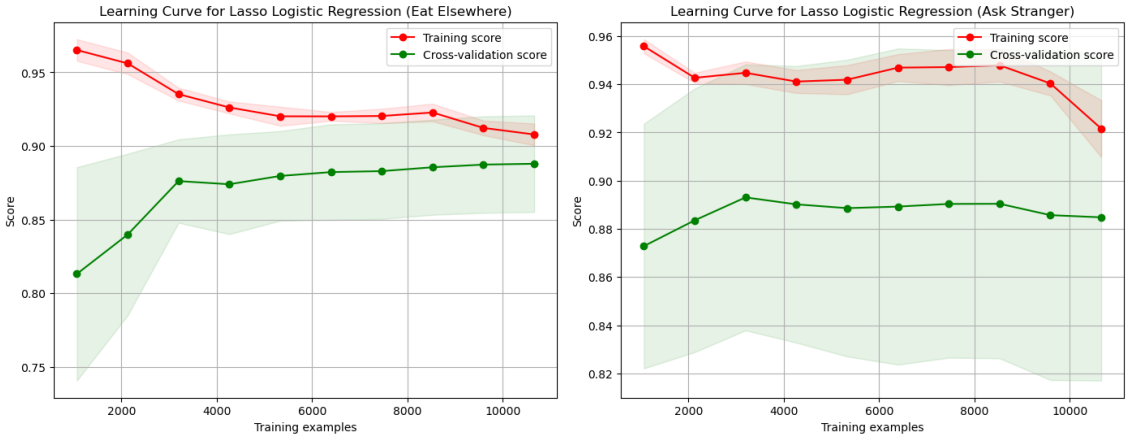
(c) (d)

Figure C.2: Learning Curves for Borrowed Food and Degrade Income Source



(e) (f)

Figure C.3: Learning Curves for Reduce Education Exp. and Sell HH Assets



(g) (h)

Figure C.4: Learning Curves for Eat Elsewhere and Ask Stranger

Bibliography

[1] OCHA. "Ukraine." (2024), [Online]. Available: https://www.unocha.org/ukraine.

[2] I. O. for Migration, "Internal displacement report: Ukraine," 2023.

[3] ACAPS, "Impact of conflict on the ukrainian economy: Implications for humanitarian needs," 2023.

[4] UNICEF. "Education under attack." (2023), [Online]. Available: https://www.unicef.org/education-under-attack.

[5] A. Vidhya. "Russia-ukraine international armed conflict: Immense damage to essential infrastructure will cause major suffering as winter looms." (2022), [Online]. Available: https://www.icrc.org/en/document/russia-ukraine-international-armed-conflict-immense-damage-essential-infrastructure.

[6] "War and conflict," DOCTORS WITHOUT BORDERS. (May 2023), [Online]. Available: https://www.doctorswithoutborders.ca/war-and-conflict/.

[7] UNICEF. "War has hampered education for 5.3 million children in ukraine, warns unicef." (2023), [Online]. Available: https://www.unicef.org/ukraine/en/press-releases/war-has-hampered-education.

[8] S. Folkman and J. T. Moskowitz, "Coping: Pitfalls and promise," *Annu. Rev. Psychol.*, vol. 55, pp. 745–774, 2004.

[9] E. C. Schoenmakers, T. G. Van Tilburg, and T. Fokkema, "Problem-focused and emotion-focused coping options and loneliness: How are they related?" *European journal of ageing*, vol. 12, pp. 153–161, 2015.

[10] E. B. Algorani and V. Gupta, "Coping mechanisms," in *StatPearls [Internet]*, StatPearls Publishing, 2023.

[11] UNHCR, "Global trends: Forced displacement in 2022," 2022.

[12] ACAPS, "Yemen life goes on in yemen conversations with yemeni families as the war nears its eighth year," 2022.

[13] D. Maxwell and R. Caldwell, "The coping strategies index: Field methods manual," *Cooperative for Assistance and Relief Everywhere Inc. USA*, 2008.

[14] UNHCR, "Coping strategies," 2020.

[15] FAOUN. "Syrian arab republic situation report." (Sep. 2019), [Online]. Available: https://reliefweb.int/report/syrian-arab-republic/syrian-arab-republic-situation-report-september-2019.

[16] M. Verpoorten, "Household coping in war-and peacetime: Cattle sales in rwanda, 1991–2001," *Journal of development Economics*, vol. 88, no. 1, pp. 67–86, 2009.

[17] J. Palma and C. Araos, "Household coping strategies during the covid-19 pandemic in chile," *Frontiers in Sociology*, p. 162, 2021.

[18] A. Ekzayez, Y. Alhaj Ahmad, H. Alhaleb, and F. Checchi, "The impact of armed conflict on utilisation of health services in north-west syria: An observational study," *Conflict and health*, vol. 15, no. 1, pp. 1–10, 2021.

[19] F. Prevatt and F. D. Kelly, "Dropping out of school: A review of intervention programs," *Journal of school psychology*, vol. 41, no. 5, pp. 377–395, 2003.

[20] R. Akresh, *Armed conflict and schooling: Evidence from the 1994 Rwandan genocide*. World Bank Publications, 2008, vol. 3516.

[21] L. B. Deng, "Are non-poor households always less vulnerable? the case of households exposed to protracted civil war in southern sudan," *Disasters*, vol. 32, no. 3, pp. 377–398, 2008.

[22] R. Hill, E. Skoufias, and B. Maher, "The chronology of a disaster: A review and assessment of the value of acting early on household welfare," 2019.

[23] P. Justino, "The impact of armed civil conflict on household welfare and policy responses," *Securing Peace: State-Building and Economic Development in Post-Conflict Countries*, pp. 19–52, 2011.

[24] M. M. Lokshin and R. Yemtsov, "Household strategies of coping with shocks in post-crisis russia," *Review of Development Economics*, vol. 8, no. 1, pp. 15–32, 2004.

[25] M. Munas and G. Lokuge, "Shocks and coping strategies of coastal communities in war–conflict-affected areas of the north and east of sri lanka," *Regional environmental change*, vol. 16, no. 2, pp. 289–299, 2016.

[26] S. Tongruksawattana, H. Waibel, and E. Schmidt, "How do rural households cope with shocks? evidence from northeast thailand," 2010.

[27] E. J. Taylor, "The new economics of labour migration and the role of remittances in the migration process," *International migration*, vol. 37, no. 1, pp. 63–88, 1999.

[28] UNHCR, "Cash programmes - lebanon," 2023.

[29] H. R. Watch, "Landmine use in ukraine," 2023.

[30] REACH, "Ukraine: Evacuations from eastern oblasts," 2022.

[31] A. Namasivayam, P. A. González, R. C. Delgado, and P. C. Chi, "The effect of armed conflict on the utilization of maternal health services in uganda: A population-based study," *PLoS currents*, vol. 9, 2017.

[32] REACH, "Ukraine: Health sector needs assessment," 2023.

[33] UNICEF, "Humanitarian situation: Yemen," 2023.

[34] WFP, "Ukraine food security trend analysis," 2023.

[35] R. J. McAlexander and L. Mentch, "Predictive inference with random forests: A new perspective on classical analyses," *Research & Politics*, vol. 7, no. 1, p. 2 053 168 020 905 487, 2020.

[36] R. Hoekstra, H. A. Kiers, and A. Johnson, "Are assumptions of well-known statistical techniques checked, and why (not)?" *Frontiers in psychology*, vol. 3, p. 137, 2012.

[37] H. S. R. Rajula, G. Verlato, M. Manchia, N. Antonucci, and V. Fanos, "Comparison of conventional statistical methods with machine learning in medicine: Diagnosis, drug development, and treatment," *Medicina*, vol. 56, no. 9, p. 455, 2020.

[38] C. Ley, R. K. Martin, A. Pareek, A. Groll, R. Seil, and T. Tischer, "Machine learning and conventional statistics: Making sense of the differences," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 30, no. 3, pp. 753–757, 2022.

[39] J. Bodenstab. "Traditional statistics versus machine learning. what's the difference?" (Sep. 2017), [Online]. Available: https://www.toolsgroup.com/blog/traditional-statistics-versus-machine-learning-whats-the-difference/.

[40] Y. Zhao, Z. S.-Y. Wong, and K. L. Tsui, "A framework of rebalancing imbalanced healthcare data for rare events' classification: A case of look-alike sound-alike mix-up incident detection," *Journal of healthcare engineering*, vol. 2018, 2018.

[41] T. A. Prout, S. Zilcha-Mano, K. Aafjes-van Doorn, *et al.*, "Identifying predictors of psychological distress during covid-19: A machine learning approach," *Frontiers in Psychology*, vol. 11, p. 586 202, 2020.

[42] A. Garbero and M. Letta, "Predicting household resilience with machine learning: Preliminary cross-country tests," *Empirical Economics*, vol. 63, no. 4, pp. 2057–2070, 2022.

[43] M. D'Angeli and P. Vesco, "Predicting the impact of armed conflict on vulnerability: A machine learning approach," in *German Institute of Development and Sustainability (IDOS), Bonn, Germany (24-25 October 2022)*, 2022.

[44] REACH. "Who we are." (2023), [Online]. Available: https://www.reach-initiative.org/who-we-are/.

[45] OCHA, "Reach initiative.," 2023.

[46] WFP. "Ukraine needs assessment: Food security and essential needs." (2023), [Online]. Available: https://docs.wfp.org/api/documents/WFP-0000150531/download/.

[47] Google. "Imbalanced data." (2023), [Online]. Available: https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data.

[48] edX. "What is undersqampling?" (2022), [Online]. Available: https://www.mastersindatascience.org/learning/statistics-data-science/undersampling/.

[49] Reuters. "Maps: Tracking the russian invasion of ukraine." (2024), [Online]. Available: https://www.reuters.com/graphics/UKRAINE-CRISIS/zdpxokdxzvx/embeds/en/map/.

[50] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in python, journal of machine learning re-search, 12," 2011.

[51] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996.

[52] F. Shofiyah and A. Sofro, "Split and conquer method in penalized logistic regression with lasso (application on credit scoring data)," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1108, 2018, p. 012 107.

[53] S. Ö. Arik and T. Pfister, "Tabnet: Attentive interpretable tabular learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, 2021, pp. 6679–6687.

[54] A. Martins and R. Astudillo, "From softmax to sparsemax: A sparse model of attention and multi-label classification," in *International conference on machine learning*, PMLR, 2016, pp. 1614–1623.

[55] D. Berrar *et al.*, *Cross-validation.* 2019.

[56] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011, https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.learning_curve.html.

[57] OpenstreetMaps. "Maps of ukraine and gis data." (2023), [Online]. Available: https://osmtoday.com/europe/ukraine.html.

[58] GeoPandas. "Introduction to geopandas." (2013–2022), [Online]. Available: https://geopandas.org/en/stable/getting_started/introduction.html.