ML for a good cause - predicting food insecurity in Chad with machine learning methods

Team 23

Abstract

Food insecurity is a serious global problem that affects and kills millions of people every year. Estimating the scale of the problem is essential to raise awareness, but this can be challenging due to the complexity of the causes and the lack of data. Recent advances in data collection and machine learning capabilities hold promise for predicting food insecurity at detailed geographic scales and frequencies. However, there is no consensus on best practices for data selection and application of methods for food insecurity modelling. Decision makers need to understand how the model works. The available data is not very large, so there is a risk that the model will overfit when the distribution changes and not be able to predict well. In this paper, we address these issues by analyzing the underlying variables and applying a set of machine learning methods to food insecurity data in 23 different regions of Chad. We showed how modeling decisions can affect the model results and analyze the impact of specific characteristics on the probability of being assigned to one of the food insecurity classes. We prove that machine learning-based methods are able to provide accurate and reliable predictions of food insecurity, the level of gender inequality has a significant positive relationship with food insecurity, intensive protectionism measures have a significant negative effect on predicted food insecurity, and for some regions the model performs better than others, and therefore it is necessary to determine which regions should be more closely monitored in terms of food insecurity.

1 Introduction

Food insecurity (FI) is a major worldwide problem, leading to suffering and death of millions of people yearly [1]. Predicting the scale of such problem is crucial for an effective response to combat hunger. However, this can be challenging due to the complexity of causes and lack of data. To address this problem, recent outbreak in data gathering capabilities and machine learning offer the promise of predicting FI at a granular geographic scale and frequency by combining geographic, weather and socioeconomic data with the prices of food. Researchers have recently started applying machine learning and statistical methods to identify FI, whereas policymakers are increasingly interested in this approach. However, turning data into forecasts is a complex process involving many small, but important decisions that affect the accuracy and usefulness of models for decision-makers.

There is still no consensus on best practice in data selection and application of the method to FI modelling. Model assumptions can have a significant impact on the results, whether it should be a standard, econometric approach that is easy to explain or a more novel, machine learning, but black-box solution that can obscure the links between modelling decisions and predictions. Decision makers, who will benefit with the implementation of the model, do not have to always understand the complexity of the model, but a vocal explanation of what drives the model, would be necessary to change their minds, to put new policies to life. This can also introduce many wrongdoings to the model performance, for example difference in distributions between training and testing samples. For example, a model based on urban data may overlook major food crises in remote, rural areas. Therefore, to develop effective FI prediction models, it is necessary to understand how the model works and what decisions it makes and why. Last but not least, the available data is usually not very large, so there is a risk of over-fitting a model that cannot make good predictions when the distribution changes. For example, a model that has been trained perfectly on only 2022 data will perform poorly in 2023.

In this paper, we address these challenges, by first analyzing the main drivers, factors that can cause FI, and only then applying selected features to solve the problem. We apply a set of machine learning algorithms to the data on FI in 23 different subregions of Chad. We show how our modelling decisions could impact the model results and we analyze the impact of particular features on the probability of falling within one of the food insecurity classes.

Our hypotheses are:

1. Machine learning based methods are able to provide precise and robust predictions on the FI measured as the phase class in a given administrative region in Chad.
2. Gender inequality levels have a significant, positive relationship with the FI measured as a phase class in a given administrative region in Chad.
3. Intensive protectionist measures, especially tariffs on food imports have a significant negative effect on the predicted FI measured as a phase class in a given administrative region in Chad.
4. For some regions model will perform better than for other, thus it is necessary to define, which regions should be monitored more carefully in regard to FI.

2 Literature review

The main focus of this research is on finding the main root causes that contribute to food insecurity. Many researchers have tackled this problem before us, using both statistical and machine learning models in an attempt to identify interactions between causal factors [2–7]. There are some applications that focus on implementing data from early warning systems, or climatogical data which is based on on predicting hazards such as drought or flooding, but because their analysis is mostly univariate, they ignore interactions with other shocks or drivers of FI, especially food prices, which significantly affect households' access to food [8, 9].

The only problem that arises is the small number of observations, which prevents an accurate analysis of the sensitivity of the model and usually leads to problems such as overfitting and underfitting, in most cases excluding the possibility of using more advanced methods, including deep learning. Therefore, when forecasting FI, we should focus on seeing if we can capture such drivers as high prices, policy errors, supply chain disruptions, or even implications of protectionist policies on the target, as well as global situation reflected by stock prices. We should also anticipate that the relationship between prices, or socio-economic variables, is non-linear, so small changes in food prices can be catastrophic for vulnerable households. Besides prices, there are also political implications on FI, such as gender inequality, conflicts or the level of democracy in the country, which can have a negative impact on FI [10–12].

Besides that, the model itself should be interpretable, so that policymakers can monitor the individual factors driving the problem. Although some techniques, such as neural networks, may not generate direct interpretations of coefficients or the explanatory variables' importance, model-agnostic interpretation tools can be used to uncover the model's prediction drivers [13, 14]. Such tools provide modelers and stakeholders with confidence that the model is picking up reasonable relationships, indicating areas of possible intervention. Interpretation is also essential for model evaluation, and error analysis must be done to determine whether the model is predicting something that does not require a model, like households or locations that are always food insecure.

Last but not least, to ensure the policymakers and stakeholders, the model has to be thoroughly tested and evaluated. Accuracy and error analysis is essential in identifying the regions that are not that well understood by the model. If the model systematically underperforms in certain regions, it should be supplemented with additional information or not used at all. Another application of error analysis is to evaluate a model while accounting for the role of food aid, food assistance, or other interventions. If the model accurately predicts food-insecure locations that subsequently receive food aid, the crisis's severity should be reduced in those areas [15].

3 Data and Transformations

The data that we used in this study comes from the The Humanitarian Data Exchange dataset for Chad country, enriched with the data from the publicly available VDEM dataset [16]. The target variable is the phase class - the level of food insecurity - in the country Chad, encoded as:

- phase 1 $\rightarrow$ minimal level,
- phase 2 $\rightarrow$ stressed level,
- phases from 3 to 5 $\rightarrow$ crisis, emergency and famine level.

All the variables are encoded on one administrative level: admin1. There are 22 unique admin1 regions in Chad. The data starts from January 2017 to December 2021.

Fig. 1: Histogram of target variable distribution throughout the years

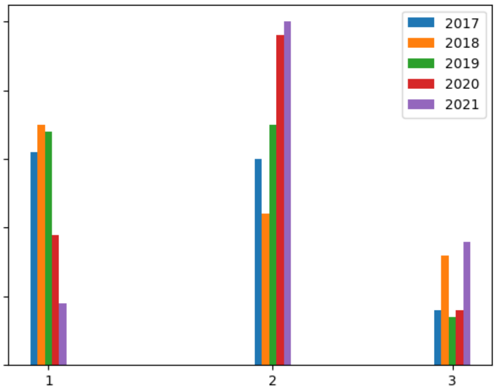


Table 1: Mean and mode for Chad phase levels throughout the years

reference_year	Mean	Mode
2017	1.67	1
2018	1.74	1
2019	1.65	2
2020	1.85	2
2021	2.12	2

Based on figure 1 and table 1, we can clearly notice that throughout the years the overall phase level has increased from 1.67 mean in 2017 to 2.12 in 2021. It might be caused by COVID pandemic crisis and the war in the Ukraine. However there are few regions that have lower mean compared to previous years.

We started our data preprocessing with handling missing observations. We have excluded every variable that had fewer than 50% of observations. We proposed a novel approach to fill missing values, by calculating the mean (mode for categorical variables) of geographically surrounding regions e.g. for Waki Fira the mean is calculated from Ennedi Ouest, Ennedi Est, Borkou, Batha, Ouaddai.

We have expanded the feature set by firstly adding new data and then performing transformations mentioned in table 2.

New variables that were added include: democratization level, gender inequality, fetched from VDEM dataset [16]. We also used more granular data from The Humanitarian Data Exchange for food prices for Chad on admin level 2. We used this data to generate minimal, mean and maximum prices of food for corresponding admin level 1 region. To cover global economy impact on food sector S&P 500 prices were added to our data. S&P 500 is an American Stock Index that includes 500 companies with highest capitalization in US. It can be used as a reflection of global economy level. We added *COVID* variable, which has binary values reflecting pandemic time in Chad from March 2020 to December 2021. We also created new features by interacting already prepared features in the data. We created *temp_amp*, which is a difference between maximum and minimum temperature in the given period. Another variable that has been created is the rainfall intensity *percip_intensity* calculated by dividing the rainfall in a given period by the number of days. Lastly, we wanted to measure import barriers and protectionism that can be impacted by government policies by subsidies and import duties. For this case, we created new variable, *import_premium_on_rice*, which is a difference between rice local prices and rice imported prices.

For every model training and evaluation, the dataset has been divided into training and testing samples and the missing for both training and testing were imputed using mean from training data.

Finally Spearman correlation, mutual information and Boruta algorithm [17] have been applied to measure the impact of explanatory variables on the target variable. The results are also presented in table 2, where feature selection is ordered from the most impactful column to the least. After data cleaning and preprocessing, we were left out with 371 observations.

We applied lagging transformation on following columns: *max_temp*, *min_temp*, *total_precip_mm*, *max_daily_precip_mm*, *rain_days*. We added lagged target variable to cover autoregressive factors. Lastly we embedded categorical variables using CatBoost encoder.

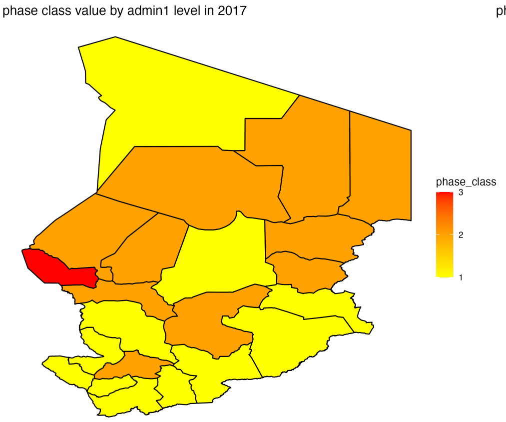
We tried to lower the dimensionality of the data using PCA algorithm, creating one variable from each group:

1. weather
2. political
3. food_prices
4. global economy

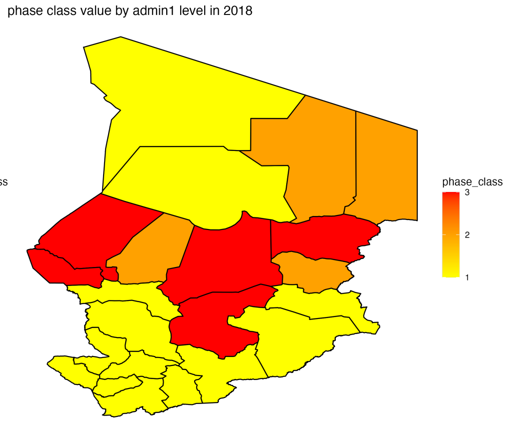
However, the final variables significantly decreased model predictive power on the training data and for that reason we decided to not use variables created by PCA algorithm.

After data cleaning we can have one last look onto the geographical distribution of the target variable in the observed years in the figure 2. We can notice large increase in the northern parts of Chad, where the climate is more desert-based.

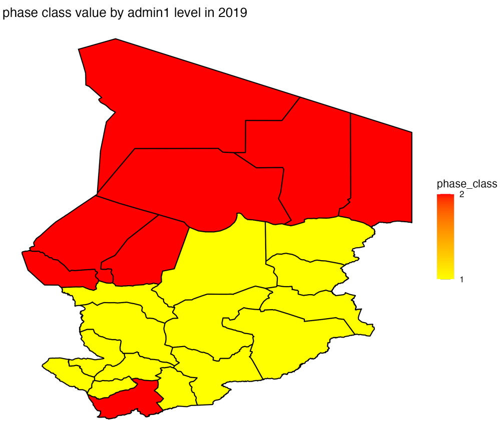
Fig. 2: Geographical distribution of target variable in the studied region. Darker colors indicate higher values.



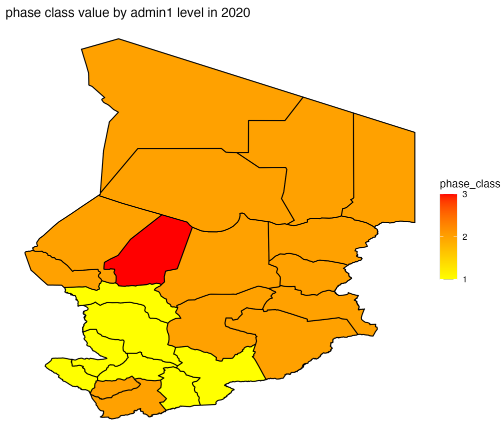
(a) target in 2017



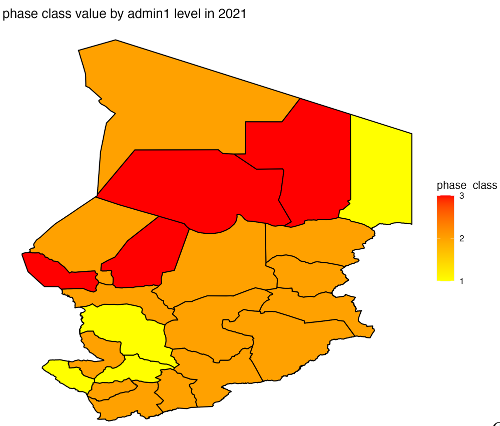
(b) target in 2018



(c) target in 2019



(d) target in 2020



(e) target in 2021

Table 2: List of variables used in the study with their feature importance index and transformations

	original variable	added feature	difference	lag	min, max, mean	embedding	feature importance
timestamp	x						
admin1	x						
phase_class	x						
reference_year	x						
max_temp	x			x	x		7
min_temp	x			x	x		8
total_precip.mm	x			x	x		17
max_daily_precip.mm	x			x			5
rain_days	x			x			2
max_sustained_wind.kph	x			x			6
duration_of_drought_months	x			x			10
flood	x					x	12
duration_of_flood_months	x			x			15
conflict	x					x	14
population	x						1
milet	x		x				4
sorghum	x		x				3
rice_local	x		x				18
rice_imported	x		x				11
maize	x		x				13
temp_amp		x					19
percip_intensity		x					
import_premium_on_rice		x					
sp500_high		x					21
sp500_low		x					
covid		x					20
gender_inequality_index		x					9

4 Models

The problem of predicting food insecurity is complicated and it is hard to assume a priori decision boundaries for changing the level of possible food shortage. Thus we made a broad and exhaustive research of different econometric and machine learning models. In this section we present a brief description of the models we used in this research, our approach to the modelling and performance criteria.

4.1 Algorithms

4.1.1 Naive approaches

Naive approaches do not assume usage of any statistical models. We created benchmark results to be able to compare our results in the most objective way. First naive approach assumes just random class drawing for every prediction, with estimated accuracy being 25%. Another two approaches assume taking results from the preceding year - first simply per administrative region, second using mode of the level of food insecurity for Chad for a specific year.

4.1.2 Ordered logit/probit

We used two types of ordinal regression - the ordered logit and ordered probit model. Ordinal regression is a statistical model used to analyze and predict the results of ordinal dependent variables. It is commonly used if a dependent variable has ordered levels – in our case "low", "medium", "high" and "very high". The model estimates the probability of an event occurring in one of the categories, given a set of independent variables. Ordinal regression assumes that the levels of the target variable are ordered in the way that the distance between the categories is equal. From the output of the regression we can obtain estimates of the coefficients for each independent variable, as well as the odds ratios and confidence intervals. Ordinal regression to model FI was used for example by [18] to estimate national food security levels in post conflict South Sudan.

4.1.3 Multinomial logistic regression

Multinomial logistic regression is an approach that estimates the choice of a categorical outcome variable, which can't be ordered in a logical way e.g. colours. It is also used for data, when the distance between the levels is unknown – we don't have the specific distance between burden levels, which is a key factor for which we have estimated this model. Multinomial logistic regression has been used by [19] to identify factors associated with the presence and severity of food insecurity among a sample of Honduran caregivers of young children.

4.1.4 Random forest

Random forest is a first of a set of machine learning algorithms used in this study. It employs an ensemble of decision trees to make predictions. It works by creating multiple decision trees on randomly selected subsets of the data and then combining their predictions to make a final prediction. This helps to reduce over-fitting and increase the accuracy of the model. Random forest can be used for both classification and regression tasks and is a popular algorithm for solving complex problems. Random forest has been also used to estimate FI, as well as it is well acknowledged in various fields such as finance, healthcare, and marketing, specifically in medical studies [20–22]

4.1.5 Naive Bayes

Naive Bayes methods are a set of supervised learning algorithms that are based on Bayes theorem with the "naive" assumption of conditional independence between every pair of variables given the value of the target variable. The main difference between naive Bayes classifiers is the assumption made regarding the distribution. In our research we chose The Gaussian Naive Bayes algorithm, which is used for classification, where the likelihood of the features is assumed to be Gaussian. Naive Bayes was used by [23] to classify FAO status of a households.

4.1.6 Support Vector Machines

The Support Vector Machines (SVM) is one of the fundamental non-parametric machine learning algorithms. The main author of this model is Professor Vladimir Vapnik [24]. The general idea of SVM is as follows: in a multi-dimensional space there exists a hyperplane which separates the classes in optimal way. The goal of SVM is to find the hyperplane, which maximizes the minimum distance (margin) between this hyperplane and observations from both classes. SVM for food insecurity estimation problem has been used for example by [25] to predict the food security in regions of Brazil.

4.1.7 K-nearest neighbours

The K-nearest neighbours (KNN) algorithm is a basic and probably the simplest supervised machine learning algorithm for both classification and regression problems. Its based on a simple idea - the best prediction for a certain observation is the known target value (label) for the observation from the training set that is most similar to the observation for which we are predicting. KNN's main advantage is that it's non-parametric (it does not require the assumption of a sample distribution) and instance-based (it does not carry out the learning process directly - it remembers the training set and creates predictions on the basis of it on an ongoing basis). It was also used to classify regions that can be food insecure based on tweets or survey data [23, 26].

4.1.8 CatBoost, LightGBM, NGBoost

CatBoost is a machine learning algorithm for gradient boosting on decision trees. The algorithm was constructed in the way it is able to handle missing values, feature interactions, and high-dimensional data. Another approach that we used was NGBoost, which also is a boosting algorithm that uses natural gradient descent to optimize the model. The difference between the two is that CatBoost estimates the expected probability for every sample, while NGBoost is able to provide a probability distribution for the whole population, due to usage of Kullback-Leiber deviation. Last algorithm is LightGBM, also a gradient boosting algorithm, but it's using several techniques (including histogram sampling) to produce the boosting results much quicker. All these the algorithms are well-recognized in the machine learning state-of-the-art [27–29].

4.1.9 AutoML

To reduce the time to provide reliable machine learning model, we also used AutoML approaches. They are based on the preparation of many, automated models providing an ensemble of the best predictors. PyCaret [30] is an open-source, low-code machine learning library that provides AutoML models, another one is MLJar [31]. Since we do not have many data points we do not focus on usage of neural networks, argumenting that they need many samples to provide reliable estimator [32]. Therefore best algorithms from AutoML packages were supposed to provide boosting, ensemble methods.

Since our training procedure includes rolling window approach, we are trying to find the best model multiple times. Therefore in every training year, the best model might be different, always providing the best possible algorithm, instead of focusing only on one family of models.

4.1.10 N-HiTS/NBEATS

Even though we stated that neural networks might not be the ideal estimator in state of low number of training observations [32], we wanted to test recurrent neural network approach, to accommodate for the panel type of data. Neural Hierarchical Interpolation for Time Series Forecasting (N-HiTS) [33] and Neural Basis Expansion Analysis for interpretable Time Series forecasting (NBEATS) [34] are two time series focused models that can accommodate for different groups of data while modelling. These architectures were thoroughly tested on several well-known datasets, including M3, M4 competition datasets, reaching very high accuracy.

4.2 Modelling approach

The prepared modeling methodology results from both: (a) modeling goals - the task of predicting categorical target variable (multiclass classification problem) for next year and (b) type of data we are dealing with - short panel data (as to time dimension – just 5 available years). Accordingly, we developed the following evaluation strategy:

1. We select appropriate econometric and machine learning model architectures that are best suited to the given research problem (based on literature and expert knowledge).
2. Using a rolling window approach we select initial 8 periods (2+ years) as initial training sample, and then we expanded the window by iteratively $i \in (0, 5)$ adding next periods.
   - training dataset (period = 1 - (8 + i))
   - testing dataset (period = 9 + i)

   Since the dataset is imbalanced, we increase weights of the smaller classes to balance it (class 1 - 35%, class 2 - 50%, class 3 -15%). Please notice that this approach insures us against the risk of data leakage and potential overfitting in case of temporal dimension.

3. Per each model architecture we perform the following operations on the training set:
   - preparation of variables specific to a particular model architecture and point in time
   - data preparation, inculding NAs imputation and feature engineering
   - model fitting
   - model stability validation using cross-validation
   - model hyperparameters tuning using cross-validation

   The final product of step 2 are 6 (corresponding to rolling window rolls) models that are finally tested on the last out-of-time available period (Sep-Dec 2021). For cross validation, we decided to use the Shuffle Split Cross Validation approach (the number of splits is 10), which allows us to introduce high randomization during the procedure. This choice seems to be a good trade-off between aware data leakage and potential overfitting within a single training sample.
4. Per each model architecture we perform following steps:
   - we create prediction on testing dataset with models trained in step 2
   - we score our testing dataset results with our novel scoring methodology
   - we perform model inspections for the impact of individual variables on the outcome using Explainable AI methods

   Please notice that we treat the test set as an out-of-sample data, to which we have no access until the final inference.

4.3 Performance criteria

The problem we are addressing is not a typical business problem - human lives may depend on the quality of our forecast, so the selection of evaluation metrics had to be done responsibly and meticulously. In our opinion, we should primarily maximize the number of observations that have been classified positively and any overclassing in the forecast will not be a big burden for us. Willing to apply our models to government recommendations, we prefer to make the mistake of overestimating the level of the burden target variable rather than underestimating it - an overestimation may result in the introduction of stronger regional policies but should not have as negative an effect on the people of that region (analogy to models indicating the spread of the COVID-19 pandemic). In contrast, underestimating the problem could lead to starvation and the death of millions of lives.

Taking this into account, we decided to approach the evaluation problem with a two-pronged approach:

- using common-known metrics:
  - Weigted F1-score,
  - Weigted Precision & Recall,
- deriving our own evaluation metrics:
  - driven underestimation punished accuracy (DUPA) – see Algorithm 4.3.1,
  - driven underestimation punished f-score (DUPF1) – see Algorithm 4.3.2.

4.3.1 Driven underestimation punished accuracy

Input:

- y_{true}: list or array of true values
- y_{pred}: list or array of predicted values
- $higher\_pen$: float representing penalty weight for true values greater than predicted values (default 1.0)
- $lower\_pen$: float representing penalty weight for true values less than predicted values (default 0.5)

Algorithm:

1. Calculate the number of y_{true} values less than y_{pred} and multiply by $lower\_pen$.
2. Calculate the number of y_{true} values greater than y_{pred} and multiply by $higher\_pen$.
3. Add the results of step 1 and step 2 together.
4. Multiply the sum from step 3 by -1.

4.3.2 Driven underestimation punished f-score

Input:

- y_{true}: list or array of true values
- y_{pred}: list or array of predicted values
- $beta$: float representing beta value for f-beta score calculation (default 2)
- $large\_penalty$: float representing penalty multiplier for f-beta scores below threshold (default 0.5)

Algorithm:

1. Calculate the confusion matrix using y_{true} and y_{pred} and store in variable cm.
2. Calculate the true positives and store in variable tp.
3. Calculate the false positives and store in variable fp.
4. Calculate the false negatives and store in variable fn.
5. Calculate the precision values and store in variable precision, using the formula $tp/(tp + fp + 1e - 9)$.
6. Calculate the recall values and store in variable recall, using the formula $tp/(tp + fn + 1e - 9)$.
7. Calculate the f-beta scores and store in variable f_beta, using the formula $(1 + beta^2) * precision * recall/(beta^2 * precision + recall + 1e - 9)$.
8. For each element in f_beta, if it is less than $large_penalty$, multiply it by $large\_penalty$.
9. Calculate the weights for each class and store in variable w.
10. Calculate the weighted f-beta score and store in variable $weighted\_f\_beta$

Table 3: Results of the metrics for all the models trained in the study

Model	DUPA	DUPF1	f1-score	precision	recall
	mean from expanding window cross-validation (6 periods)				
Naive model - random class	-15.071429	0.170875	0.334456	0.540056	0.26569
Naive model - martignal (values from prev. Year)	-9.428571	0.50509	0.579857	0.700359	0.570129
Naive model - mode for country (values from pre. Years)	-13.642857	0.341468	0.432739	0.534464	0.449562
Ordered probit	-22.857143	0.041056	0.067049	0.082424	0.088927
Ordered logit	-22.857143	0.041168	0.065969	0.074248	0.088927
Naive Bayes	-8.9100529	0.550563	0.573272	0.643941	0.614053
KNN	-9.1216931	0.609237	0.660852	0.715502	0.675584
SVM	-7.4312169	0.672355	0.660127	0.683306	0.713332
Random Forest	-6.0661375	0.771044	0.753463	0.738074	0.786266
CatBoost	-6.1560846	0.776415	0.785524	0.834733	0.808199
ml-jar (ensemble of XGBoost, Extra Random Tree, Random Forest)	-6.86124	0.73054	0.72651	0.731368	0.737637
PyCaret (LightGBM)	-2.928571	0.785559	0.801761	0.824765	0.826287
NBeatsX	-3.694628	0.775991	0.805584	0.808959	0.811563
NHITS	-4.852123	0.744698	0.732459	0.737546	0.750601
	out of sample				
pyCaret - LightGBM	-2	0.910973	0.911111	0.921739	0.913043

Fig. 3: DUPA score for different models

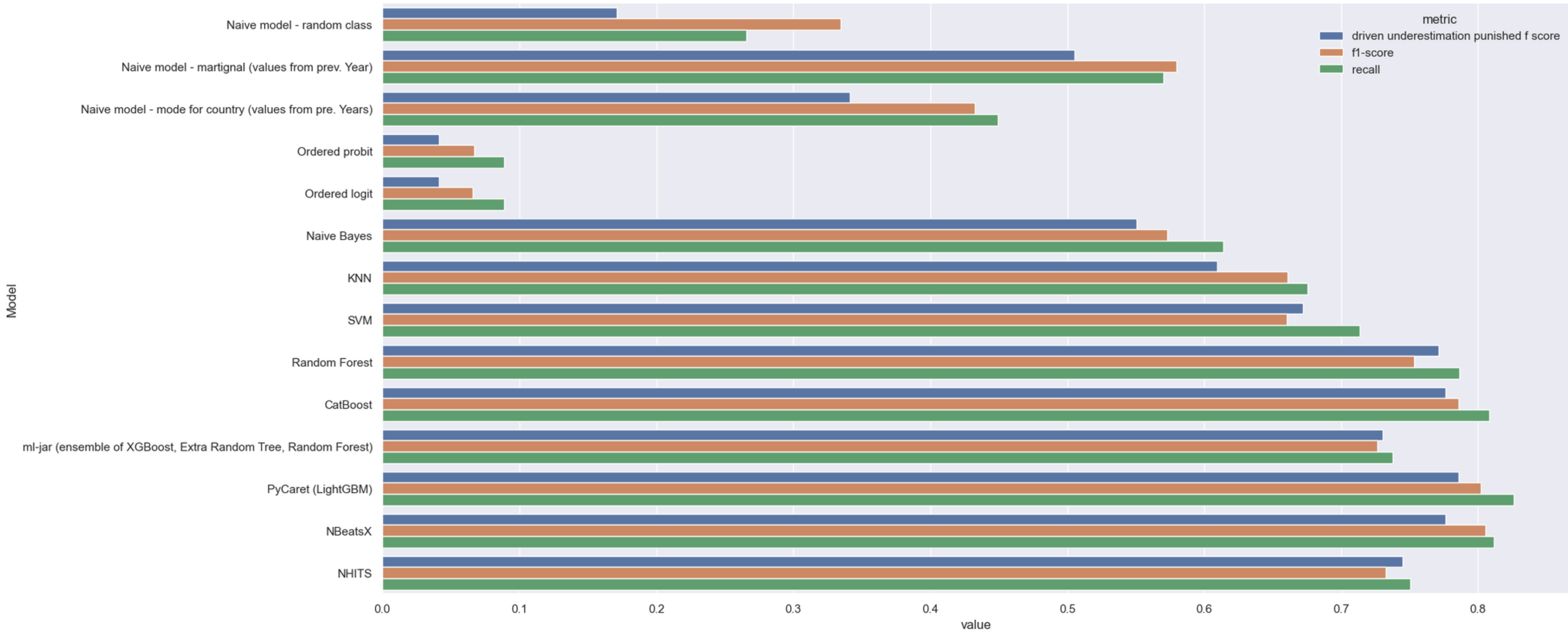
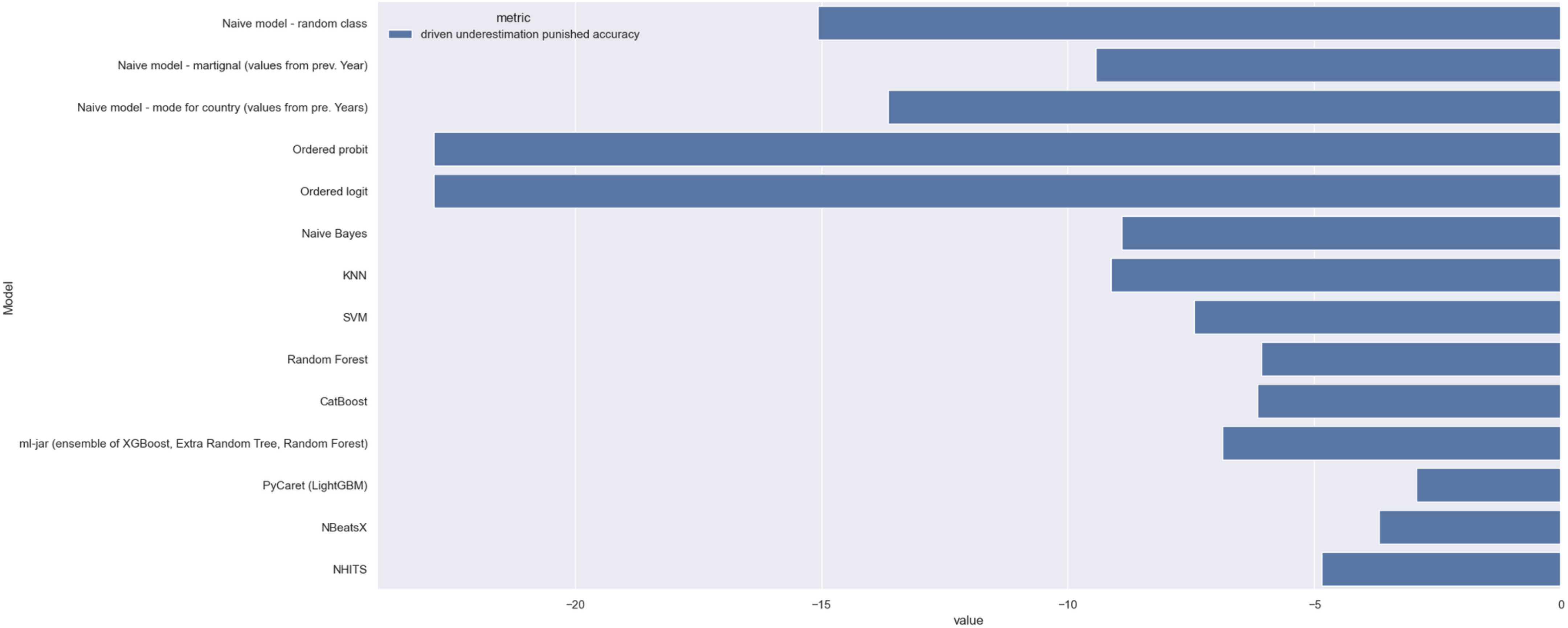


Fig. 4: F1 score for different models



5 Results

5.1 Models performance

Table 3 and figures 3, 4 contain the results of our models on (a) 6 period expanding window cross-validation and (b) out-of-sample testing period. Following the cross-validation modelling procedure (a), we selected a global champion, which turned out to be the LightGBM model (generated by AutoML PyCaret library). For this specific architecture we get the following results (average from 6 periods) on our metrics: DUPA/DUPF1/F1 in cross-validation procedure: -2.928571/ 0.785559/ 0.801761. These results seem to be very satisfactory in the context of the performance of other architectures of statistical models and machine learning. It should be noted that we managed to outperform all naive benchmark models in this way (interestingly, the best of them is martignal (values from prev. Year). Moreover, it is worth noting that the class of models that best coped with this problem were tree models based on the idea of bagging (Random Forest) and boosting (CatBoost, LightGBM, XGBoost). At the same time, we would like to point out that the second approach we proposed based on time series forecasting using the global neural network models like NHITS and NBEATS models also proved successful (this model gained the highest f1-score). In addition to PyCaret, we tried the AutoML ml-jar package, which selected the model ensemble weigted model from XGBoost, Random Forest, Extra Random Tree as the best - however, as it turned out during the test, such greedy assembling led to strong overfitting and cannot be used in a practical problem. Finally, based on our empirical results, we conclude that the LightGBM model preceded by our proprietary preprocessing is the best solution to the research problem under consideration.

Finally, using our champion model (LightGBM), we performed inference on set (b) out-of-sample. The results we obtained from this model are DUPA/DUPF1/F1-Score/Recall: -2/ 0.910/ 0.911/ 0.913. Detailed classification results are included in the 5 figure. Our model made two misclassifications, more precisely it got it wrong once for the actual class: 2 and 3. Such a result is very satisfying looking at the class of the problem under consideration.

It is worth noting that as part of the modelling, we also analysed the probability cut-off points (on the training set) for the different classes, in order to optimise the recall metric in the precision-recall tradeoff. An example of the curve we optimised is included in the figure 6.

Fig. 5: Confusion matrix for the best model

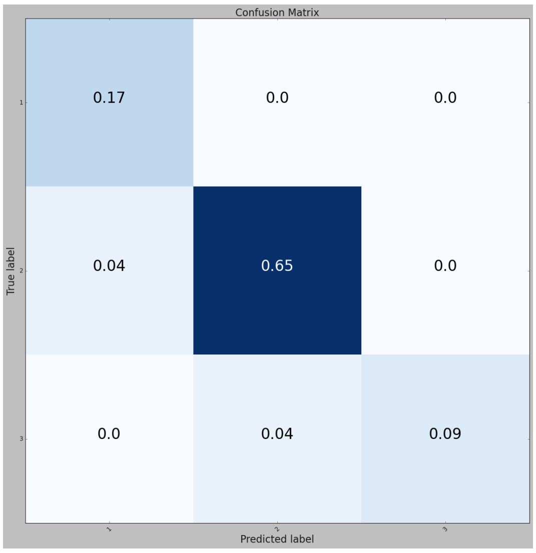
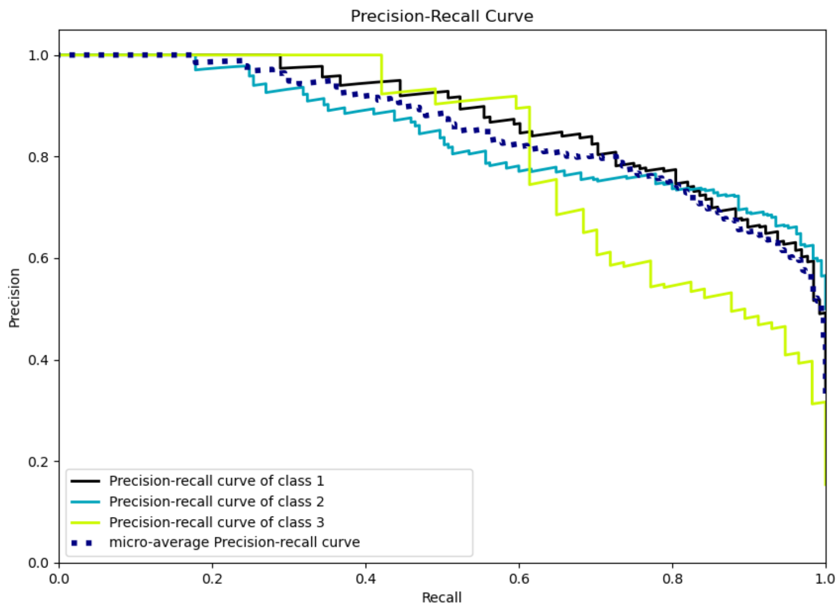


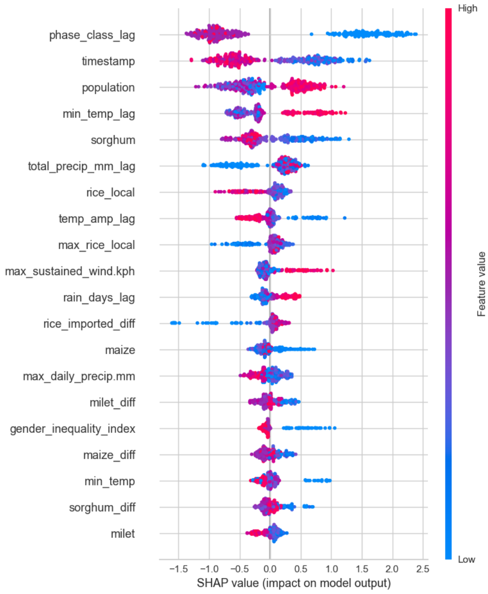
Fig. 6: Precision-recall curve for the model



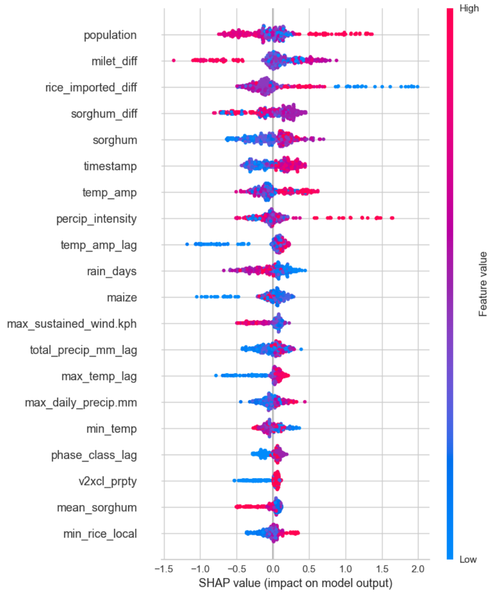
5.2 Model drivers and insights - XAI

For an ordered choice problem, the explainable AI (XAI) methods are still underdeveloped. The publicly available XAI tools do not distinguish between ordered and unordered choice problems. AI tools suggest performing the analysis with one-versus-rest approach, meaning that the choice problem is divided into n separate binary classification problems, where n is the number of label levels in the original problem. This means that the interpretation of XAI method results can be unintuitive and time consuming, especially when compared to the interpretation process in regression or binary classification problems. Nevertheless, for ordered choice problems XAI is still an excellent tool to derive meaningful insights from black box models of this paper, which is key for the shareholders, here policymakers. We verify Hypothesis 2 and Hypothesis 3, as well as derive insights about the prediction strategy of the best performing model, using the available XAI tools from the shap library in python.

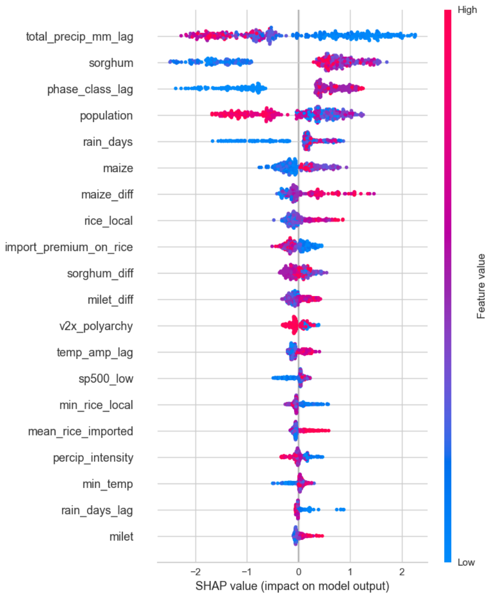
Fig. 7: SHAP values per target category



(a) target = 0



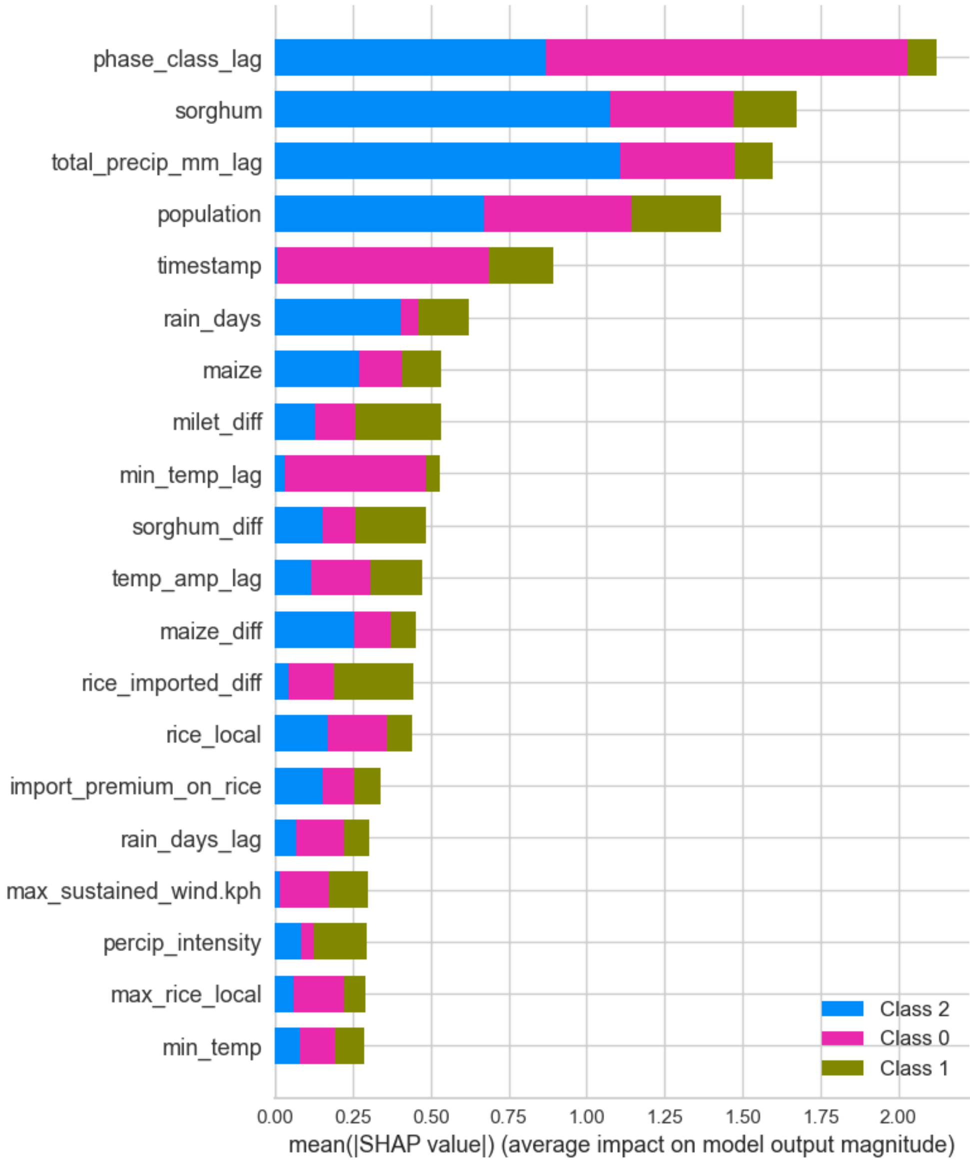
(b) target = 1



(c) target = 2

Figure 7 shows the magnitude of information to be gained from using the XAI methods combined with black box algorithms. The variables in the figures are ordered by having the variables with most significant impact on the final label value being at the top, and the ones with the least impact being at the bottom. When analyzing the plots, we find that one of the most significant predictors when considering the label = 3, which coincides with high food insecurity is sorghum, which is the price of sorghum - which constitutes almost 19% of daily caloric intake in Chad [35]. The figure shows that high prices of sorghum are associated with high probability of the model predicting high food insecurity levels. The figure for the predicted label = 1, meaning low food insecurity, shows that among the most significant drivers behind predicting this value of label are min_temp_lag - the lowest temperature measured in the period before the one used for prediction.. For minimum temperature in the preceding period, the higher the value the higher the chance of predicting low food insecurity. This might mean that crops in Chad can suffer from low temperatures, so some crops with higher low temperature tolerance could be introduced to alleviate the risk of food insecurity.

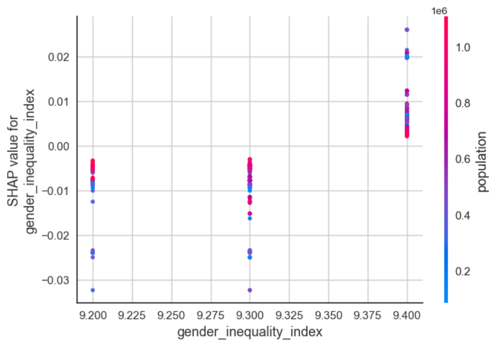
Fig. 8: SHAP values for best model without distinction



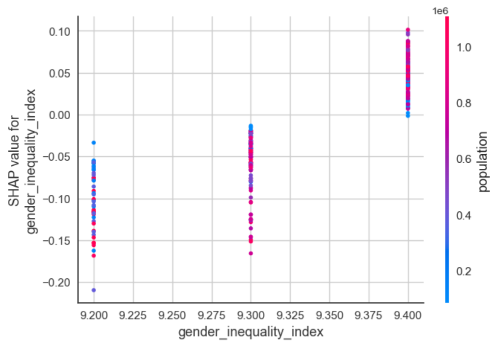
The combined SHAP value plot suggests two conclusions about the performance and predictive strategy of our best performing model. For one, of the variables with the highest impact on the model is timestamp, which is the identifier of the period that is being predicted. This suggests that this variable inherited the explanatory power of some of the variables not included in the model that are location invariant or at least consistent at the national level. This results in the variable containing the explanatory power of time-fixed effects in Chad. Such effects include the seasonal changes in food insecurity due to annual or bi-annual cycle of crop growing. The second important takeaway is that variables most impacting the predicted level of food insecurity are related to climate conditions, for example total precipitation in the previous period, and to the prices of crops which a typical diet in Chad consists of. This is in line with intuition as these seem to be the variables most closely related to the performance of the agricultural sector.

In order to focus on one or two given predictors, which is most likely the most useful for the policymakers, a partial dependence plot can be constructed. The partial dependence plot shows more precisely how the level of a given predictor affects the final prediction. In line with the shap plots, the partial dependence profiles have to be plotted with the one-versus-rest approach.

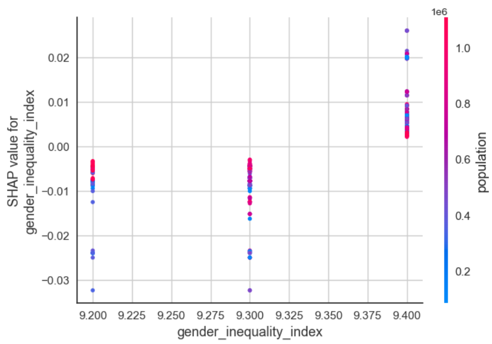
Fig. 9: Partial Dependence Plots for variable GII values in distinction for target values



(a) target = 0



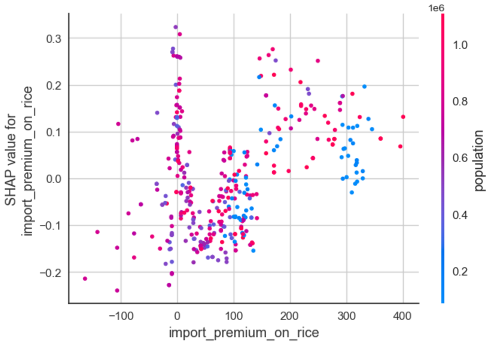
(b) target = 1



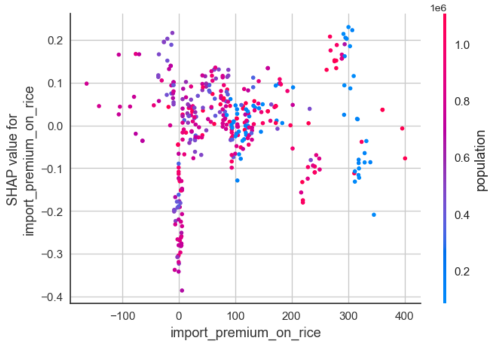
(c) target = 2

Figure 9 shows the partial dependence profiles of Gender.Inequality.Index (GII), being our stand in for female employment in this research, as the female employment data was not available for the scope of our research. The general idea underlying the construction of PD profiles is to show how the expected value of model prediction behaves as a function of a selected explanatory variable. This means that if the slope of the best line in the plot would be positive, the variable has a positive relationship with the label. With GII being the highest in regions with the most unequal treatment based on gender, the results seem not to be intuitive. The model interprets the high values of GII as being associated with medium and high levels of food insecurity. For the value of label corresponding to low food insecurity, the relationship between GII and the probability of predicting low food insecurity is positive, in line with Hypothesis 2. This means that the implied relationship between GII and food insecurity levels is, as has been widely accepted by the scientific authorities, positive. It is likely that due to limited scope of data the model used for the implied relationship captures a short term trend rather than a long term one, so it would be good to verify these results over a longer time period. GII is not a variable that changes a lot over time, so if the data was sufficient to perform a robust panel analysis, it is likely that within-region variations in GII should be significant enough to derive even more meaningful insights for the policy makers.

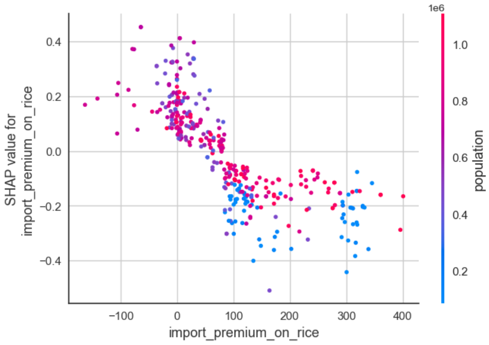
Fig. 10: Partial Dependence Plots for variable malaria_fever values in distinction for target values



(a) target = 0



(b) target = 1

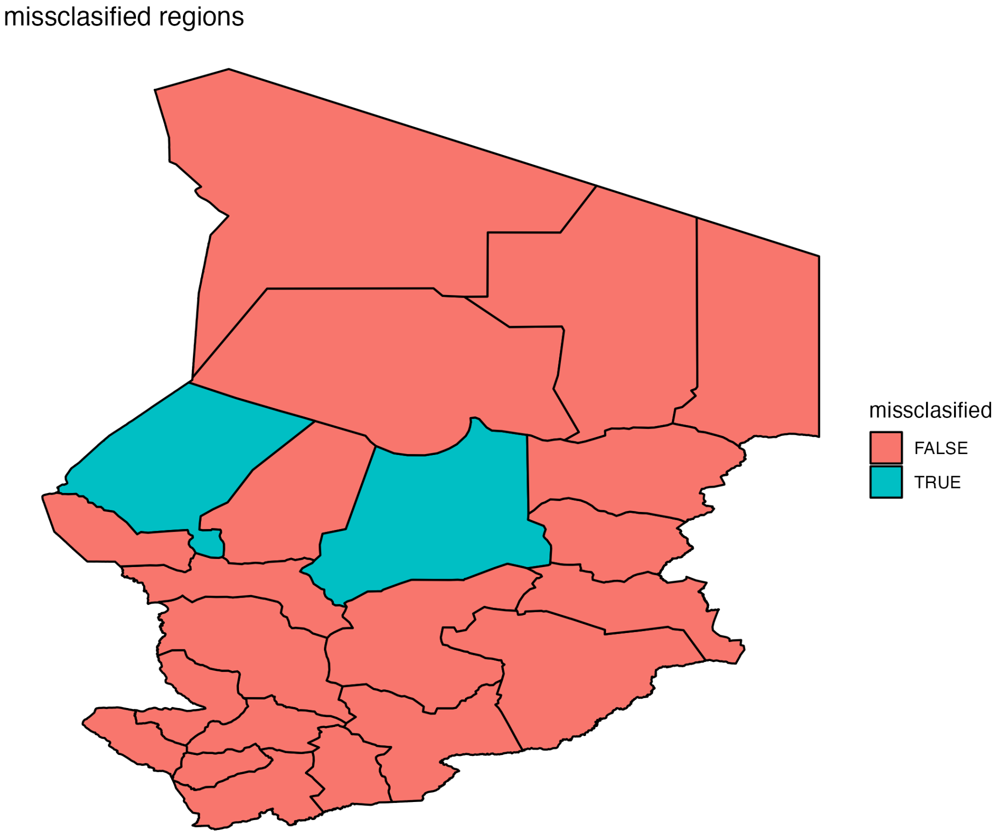


(c) target = 2

Figure 10 shows the impact of the difference in price between the imported rice and the locally grown one - import_premium_on_rice. Here, the scientific consensus is that the relationship is significant and negative, which is in line with what the partial dependence plot show. For the plot corresponding with high food insecurity, the impact of import_premium_on_rice is clearly negative, as the higher the level of protectionism towards the agricultural sector the lower the probability of predicting high food insecurity. This might not be intuitive, as it is usually the case that protectionist measures introduce market inefficiencies, however in the case of the agricultural sector it is commonly accepted that protectionist measures allow local businesses to thrive even when they are less efficient than their foreign counterparts. This insulates the market from external supply shocks in the food market. In the observed period such shocks occurred e.g the COVID-19 pandemic, which without the insulation would have greatly impacted the food insecurity in Chad.

Figure 11 shows the impact of the model's misclassification for two regions of Chad: Batha and Kanem. Relating to the literature, the problem is very crucial when it comes

Fig. 11: Best models misclassification for the out-of-time sample



to underestimation, therefore we would like to give further implications to the policy-makers that more thorough data analysis should be performed for these regions. We believe that this phenomenon could be corrected by delivering more data or providing new variables that are provided by people with domain knowledge. Otherwise, another good advice would be to put more emphasis on monitoring these regions.

6 Discussion and recommendations

This paper addressed the issue of predicting the food insecurity in different regions in Chad, over the 2017-2021 period. The analyses show that machine learning based approaches prove to be very effective at predicting food insecurity, and greatly outperform more naive approaches such as the martingal approach. We have also proposed a framework for developing models and two new model performance evaluation metrics, fit for the problem of ordered choice modeling. One of the limitations of the machine learning based approaches used to be the leak of model interpretability, however basing on the recent developments in the field of explainable AI, we have shown that even the most complex models can be used to derive meaningful and intuitive insights, that may aid the policymakers. In that vein, we have performed an analysis of the impact of gender inequality index on the predicted level of food insecurity, and have concluded that higher levels of gender inequality coincide with high probabilities of predicting high levels of food insecurity. This leads us to a conclusion that we do not have sufficient evidence to reject hypothesis 2. We have also analyzed the impact of the intensity of protectionist measures on the prediction, concluding that in line with contemporary literature, protectionist measures such as tariffs can help reduce the probability of high food insecurity. Finally we focused on checking the last hypothesis, which considers model quality for every particular region. It seems that there were two regions that were misclassified by the best model, we believe a further monitoring of available data could be performed or additional business specific data provided to help the model perform better in these regions.

In order to provide more robust insights related to this problem more data has to be collected. This would allow researchers to analyze long-term drivers of food insecurity. Furthermore, a more time-series oriented analysis, which was out of scope for this research could prove insightful by focusing specifically on the time varying effects. We also believe that there is a need to exclude an effect of already existing worldwide humanitary aid, e.g. some regions could have already obtained food subsidies, therefore decreasing the food insecurity

Further study of the food insecurity drivers is crucial to accurately assessing the welfare impacts of many public policies. This research suggests that policies propagating gender equality, and protectionist policies for the agricultural sector may be useful as tools of reducing food insecurity. Furthermore, the study shows the need for additional research of this data, and rough direction that additional research may proceed in on this topic.

References

[1] THE STATE OF FOOD SECURITY AND NUTRITION IN THE WORLD 2020. https://doi.org/10.4060/CA9692EN . https://www.fao.org/3/ca9692en/online/ca9692en.html Accessed 2023-04-21

[2] Early Warning and Early Action for Increased Resilience of Livelihoods in IGAD Region. https://fic.tufts.edu/publication-item/early-warning-and-early-action-for-increased-resilience-of-livelihoods-in-igad-region/ Accessed 2023-04-21

[3] Lentz, E.C., Michelson, H., Baylis, K., Zhou, Y.: A data-driven approach improves food insecurity crisis prediction. World Development **122**, 399–409 (2019) https://doi.org/10.1016/j.worlddev.2019.06.008 . Accessed 2023-04-21

[4] Rome Declaration and Plan of Action. https://www.fao.org/3/w3613e/w3613e00.htm Accessed 2023-04-21

[5] Barrett, C.B.: Measuring food insecurity. Science (New York, N.Y.) **327**(5967), 825–828 (2010) https://doi.org/10.1126/science.1182768

[6] Christensen, C., Wagner, T., Langhals, B.: Year-Independent Prediction of Food Insecurity Using Classical and Neural Network Machine Learning Methods. AI **2**(2), 244–260 (2021) https://doi.org/10.3390/ai2020015 . Number: 2 Publisher: Multidisciplinary Digital Publishing Institute. Accessed 2023-04-21

[7] Nica-Avram, G., Harvey, J., Smith, G., Smith, A., Goulding, J.: Identifying food insecurity in food sharing networks via machine learning. Journal of Business Research **131**, 469–484 (2021) https://doi.org/10.1016/j.jbusres.2020.09.028 . Accessed 2023-04-21

[8] Biber-Freudenberger, L., Ergeneman, C., Förster, J.J., Dietz, T., Börner, J.: Bioeconomy futures: Expectation patterns of scientists and practitioners on the sustainability of bio-based transformation. Sustainable Development **28**(5), 1220–1235 (2020). Publisher: John Wiley & Sons, Ltd. Accessed 2023-04-21

[9] Pereira, M., Oliveira, A.M.: Poverty and food insecurity may increase as the threat of COVID-19 spreads. Public Health Nutrition **23**(17), 3236–3240 (2020) https://doi.org/10.1017/S1368980020003493

[10] Knipe, P.K.: The link between gender inequality and food security among female students at tertiary institutions in South Africa (2019). Accepted: 2020-11-27T12:32:05Z Publisher: University of Western Cape. Accessed 2023-04-21

[11] Wegren, S.K., Nikulin, A., Trotsuk, I.: Food Policy and Food Security: Putting Food on the Russian Table. Lexington Books, ??? (2018). Google-Books-ID: g29RDwAAQBAJ

[12] Bala, B.K., Alias, E.F., Arshad, F.M., Noh, K.M., Hadi, A.H.A.: Modelling of food security in Malaysia. Simulation Modelling Practice and Theory **47**, 152–164 (2014) https://doi.org/10.1016/j.simpat.2014.06.001 . Accessed 2023-04-21

[13] Andree, B.P.J., Chamorro, A., Kraay, A., Spencer, P., Wang, D.: Predicting Food Crises (2020) https://doi.org/10.1596/1813-9450-9412 . Publisher: World Bank, Washington, DC. Accessed 2023-04-21

[14] Tjoa, E., Guan, C.: A Survey on Explainable Artificial Intelligence (XAI): Towards Medical XAI. IEEE Transactions on Neural Networks and Learning Systems **32**(11), 4793–4813 (2021) https://doi.org/10.1109/TNNLS.2020.3027314 . arXiv:1907.07374 [cs]. Accessed 2023-04-21

[15] Headey, D., Barrett, C.B.: Measuring development resilience in the world's poorest countries. Proceedings of the National Academy of Sciences **112**(37), 11423–11425 (2015) https://doi.org/10.1073/pnas.1512215112 . Publisher: Proceedings of the National Academy of Sciences. Accessed 2023-04-21

[16] The V-Dem Dataset – V-Dem. https://v-dem.net/data/the-v-dem-dataset/ Accessed 2023-04-20

[17] Kursa, M.B., Rudnicki, W.R.: Feature Selection with the Boruta Package. Journal of Statistical Software **36**, 1–13 (2010) https://doi.org/10.18637/jss.v036.i11 . Accessed 2023-04-20

[18] Lokosang, L.B., Ramroop, S., Hendriks, S.L.: Establishing a robust technique for monitoring and early warning of food insecurity in post-conflict South Sudan using ordinal logistic regression. Agrekon **50**(4), 101–130 (2011) https://doi.org/10.1080/03031853.2011.617902 . Publisher: Routledge _eprint: https://doi.org/10.1080/03031853.2011.617902. Accessed 2023-04-21

[19] Ben-Davies, M.E., Kinlaw, A., Campo, Y.E.d., Bentley, M.E., Siega-Riz, A.M.: Risk factors associated with the presence and severity of food insecurity in rural Honduras. Public Health Nutrition **17**(1), 5–13 (2014) https://doi.org/10.1017/S1368980013002048 . Publisher: Cambridge University Press. Accessed 2023-04-21

[20] Doreswamy, D., Nigus, M.: Feature Selection Methods for Household Food Insecurity Classification. In: 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), pp. 1–7 (2020). https://doi.org/10.1109/ICCSEA49143.2020.9132945

[21] Alam, M.Z., Rahman, M.S., Rahman, M.S.: A Random Forest based predictor for medical data classification using feature ranking. Informatics in Medicine Unlocked **15**, 100180 (2019) https://doi.org/10.1016/j.imu.2019.100180 . Accessed 2023-04-20

[22] Pauly, O.: Random Forests for Medical Applications. PhD thesis, Technische Universität München (2012). https://mediatum.ub.tum.de/1094727 Accessed 2023-04-20

[23] Razzaq, A., Ahmed, U.I., Hashim, S., Hussain, A., Qadri, S., Ullah, S., Nawaz Shah, A., Imran, A., Asghar, A.: An Automatic Determining Food Security Status: Machine Learning based Analysis of Household Survey Data. International Journal of Food Properties **24**(1), 726–736 (2021) https://doi.org/10.1080/10942912.2021.1919703 . Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10942912.2021.1919703. Accessed 2023-04-21

[24] Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning **20**(3), 273–297 (1995) https://doi.org/10.1007/BF00994018 . Accessed 2023-04-20

[25] Barbosa, R.M., Nelson, D.R.: The Use of Support Vector Machine to Analyze Food Security in a Region of Brazil. Applied Artificial Intelligence **30**(4), 318–330 (2016) https://doi.org/10.1080/08839514.2016.1169048 . Accessed 2023-04-21

[26] Lukyamuzi, A., Ngubiri, J., Okori, W.: Tracking food insecurity from tweets using data mining techniques. In: Proceedings of the 2018 International Conference on Software Engineering In Africa. SEiA '18, pp. 27–34. Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3195528.3195531 . https://dl.acm.org/doi/10.1145/3195528.3195531 Accessed 2023-04-21

[27] Duan, T., Avati, A., Ding, D.Y., Thai, K.K., Basu, S., Ng, A.Y., Schuler, A.: NGBoost: Natural Gradient Boosting for Probabilistic Prediction. arXiv. arXiv:1910.03225 [cs, stat] (2020). https://doi.org/10.48550/arXiv.1910.03225 . http://arxiv.org/abs/1910.03225 Accessed 2023-04-20

[28] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A.: CatBoost: unbiased boosting with categorical features. arXiv. arXiv:1706.09516 [cs] (2019). https://doi.org/10.48550/arXiv.1706.09516 . http://arxiv.org/abs/1706.09516 Accessed 2023-04-20

[29] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y.: LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Advances in Neural Information Processing Systems, vol. 30. Curran Associates, Inc., ??? (2017). https://proceedings.neurips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html Accessed 2023-04-21

[30] Ali, M.: PyCaret: An Open Source, Low-code Machine Learning Library in Python. (2020). PyCaret version 1.0.0. https://www.pycaret.org

[31] Płońska, A., Płoński, P.: MLJAR: State-of-the-art Automated Machine Learning Framework for Tabular Data. Version 0.10.3. MLJAR Sp. z o.o., Łapy, Poland (2021). https://github.com/mljar/mljar-supervised

[32] Goodfellow, I.J., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge, MA, USA (2016). http://www.deeplearningbook.org

[33] Challu, C., Olivares, K.G., Oreshkin, B.N., Garza, F., Mergenthaler-Canseco, M., Dubrawski, A.: N-HiTS: Neural Hierarchical Interpolation for Time Series Forecasting. arXiv. arXiv:2201.12886 [cs] (2022). https://doi.org/10.48550/arXiv.2201.12886 . http://arxiv.org/abs/2201.12886 Accessed 2023-04-21

[34] Oreshkin, B.N., Carpov, D., Chapados, N., Bengio, Y.: N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. arXiv. arXiv:1905.10437 [cs, stat] (2020). https://doi.org/10.48550/arXiv.1905.10437 . http://arxiv.org/abs/1905.10437 Accessed 2023-04-21

[35] Program, U.S.C., Lindsay, J.: Sorghum: An Ancient, Healthy and Nutritious Old World Cereal. INTSORMIL Scientific Publications (2010)