Predicting Global Food Insecurity: AI for Identifying Countries in Need of External Assistance

Juan P Alzate

Septiembre 2024
London, United Kingdom

Contents

1. Introduction

As of the latest data from United Nations World Food Programme, 783 million people worldwide continue to suffer from hunger, a staggering number that, after years of decline, is once again on the rise. Ending hunger is one of the 17 United Nations Sustainable Development Goals (SDGs) to be achieved by 2030. This goal, proposed by the United Nations in a resolution adopted in July 2017 [3], is the second of the SDGs and aims to eradicate hunger, ensure food security, improve nutrition, and promote sustainable agriculture by 2030. Despite significant efforts under this initiative, the data reveals a worrying increase in hunger since 2019, underscoring the urgency of innovative solutions to reverse this trend and keep progress on track.

The United Nations emphasises that "a world with zero hunger can positively impact our economies, health, education, equality, and social development" [16]. Addressing hunger is therefore recognised as an interdisciplinary challenge, requiring a multi-faceted approach that brings together diverse resources, knowledge, and expertise.

One promising area in the fight against hunger is the application of artificial intelligence (AI). Over the past decade, AI has rapidly gained traction as a tool capable of addressing complex problems. In this context, AI offers potential for the early detection of famine, providing decision-makers with valuable insights before crises fully emerge.

Accompanying recent technological developments, various monitoring systems have been established to aid in food security prevention and decision-making. Notable examples include the Famine Early Warning Systems Network (FEWS NET), developed and maintained by NASA, and the Global Information and Early Warning System on Food and Agriculture (GIEWS), managed by the Food and Agriculture Organization (FAO). These systems provide real-time, global-scale data, free of charge, to inform policy and intervention strategies.

While these systems play an essential role in food security monitoring, they face certain limitations. The rankings and assessments produced by these systems are composites of multiple metrics and sources, which often results in delayed outputs. This delay stems from the intricate process of gathering, standardising, and validating large amounts of data, which frequently involves manual or inferential inputs. The complexity of these operations also limits the depth of the models, leaving little room for more sophisticated analyses.

Moreover, these monitoring systems primarily reflect the present or the immediate past, rather than offering predictive insights into future conditions. They excel at capturing what is happening or what has already occurred in a given region but are less suited to forecasting medium- or long-term trends that are critical for timely interventions.

Given the complexity of food insecurity—driven by interconnected social, economic, climatic, and spatial factors—there is a growing need to explore new approaches. Predictive models, leveraging diverse and heterogeneous data sources, hold the potential to anticipate future crises and provide more proactive responses. Such models could assess the impact of various factors on food security, enabling more efficient and timely interventions.

Identifying populations at risk of famine is a challenging task that requires a comprehensive, holistic approach. Socio-economic, environmental, and climatic factors all contribute to food insecurity, and the ability to accurately predict and mitigate risks depends on integrating these diverse elements [7]. AI offers the potential to synthesise these variables into predictive models that can not only forecast food insecurity but also identify the regions and populations most at risk.

Proactive action is vital. By moving from a reactive stance—responding only when a crisis is already underway—to a proactive approach, we can reduce the number of people at risk and build resilience in vulnerable regions [41]. Such an approach will be essential for food-insecure countries in the foreseeable future as they face the ongoing challenge of hunger.

1.1. Objective and contribution

The objective of this work is to forecast which countries will require external assistance for food, providing a measure of food security. By analysing historical data on *food loss*, *infrastructure, vegetation change*, *flooding*, *armed conflict*, *land degradation*, *human development index*, and *access to potable water*, the study evaluates the predictive power of these variables. The goal is to classify risks accurately for the future, when the metrics of the current year have been released. This novel approach not only predicts outcomes with high confidence but also identifies the most influential variables for intervention, enabling optimisation of the food chain. In doing so, the number of people at risk can be minimised with a proactive, rather than reactive, approach to food insecurity.

The innovation of this work lies in its global-scale use of a diverse set of indicators to predict the Agricultural Stress Index System (ASIS) and develop an algorithm capable of forecasting future events. This is achieved by compiling, processing, transforming, and modelling current data, irrespective of the country in question.

Another key contribution of this project is its reliance on open data, namely publicly available data that can be accessed by anyone online at no cost. The use of open data ensures transparency and replicability, making this work an important resource for global efforts to address food security.

2. Literature Review

2.1. The current landscape: Famine in data

According to projections by the United Nations, 600 million people worldwide will face hunger by 2030 [35]. This alarming forecast is expected to have cascading effects on other Sustainable Development Goals (SDGs), such as Good Health and Well-being, Quality Education, and Reduced Inequalities. These implications are of deep concern not only to the FAO (Food and Agriculture Organization of the United Nations) but also to global institutions like IFAD (International Fund for Agricultural Development), UNESCO (United Nations Educational, Scientific and Cultural Organization), the World Tourism Organization, and the WHO (World Health Organization).

The current reality already reflects an ongoing crisis. In 2023, over 274 million people globally experienced acute levels of food insecurity. This burden was distributed unevenly across different regions: 49.6 million in 13 countries in Central and Southern Africa, 64.2 million in 8 countries in East Africa, 44.3 million in 14 countries in West Africa, 59.8 million in 5 countries in Asia, 19.7 million in 9 countries in Latin America and the Caribbean, and 36.7 million in 9 countries in the Middle East and North Africa [43].

Africa remains the most critically affected continent, with around 36 countries facing high levels of acute food insecurity in 2023. Historically, Africa has consistently been the most vulnerable region, both in terms of the number of countries affected and the absolute number of people suffering from food insecurity. Pattern that seems to repeat by August 2024 as seen in Table 1. The proportion of affected populations is also higher in Africa than in any other region, a trend that has persisted for decades.

The FAO's Global Information and Early Warning System (GIEWS) regularly monitors global food security and identifies countries requiring external assistance for food. The FAO defines such countries as:

> *"Countries that are expected to lack the resources to deal with reported critical problems of food insecurity."*

This metric underscores the severity of the crisis and highlights nations where domestic capacity to address food shortages is insufficient, requiring international intervention.

The data reveals a growing need for proactive strategies to address hunger, emphasising the importance of early detection systems like GIEWS. While such systems provide invaluable insights into present and recent conditions, their scope is often limited to reactive responses rather than offering forecasts that could enable preemptive action. Addressing this gap is crucial for reducing the number of countries and individuals at risk and for achieving long-term food security on a global scale.

Exceptional shortfall	Severe localised food insecurity	Widespread lack of access	No Risk of food insecurity
Central African Rep	Afghanistan	Burundi	Rest of the world
Kenya	Burkina Faso	Dem Rep of the Congo	
Somalia	Bangladesh	Djibouti	
Sudan	Cameroon	Eritrea	
Zambia	Congo	Ethiopia	
Zimbabwe	Guinea	Haiti	
	Liberia	Lebanon	
	Libya	Sri Lanka	
	Lesotho	Mauritania	
	Madagascar	Malawi	
	Mali	Niger	
	Myanmar	Nigeria	
	Mozambique	North Korea	
	Namibia	South Sudan	
	Pakistan	Chad	
	Sudan	Ukraine	
	Senegal	Venezuela	
	Sierra Leone	Yemen	
	Eswatini	Zimbabwe	
	Uni Rep of Tanzania	Palestine	
	Uganda	Syrian Arab Republic	
	Zambia		
	Haiti		
	Ukraine		

Table 1. This table categorises countries based on the level of food insecurity risk they face according to GIEWS in September 2024, divided into four not mutually exclusive categories: Exceptional Shortfall, Severe Localised Food Insecurity, Widespread Lack of Access, and No Risk of Food Insecurity.

2.2. Machine learning in forecasting

Artificial intelligence, combined with the vast amount of data and powerful computational tools available today, offers a potential solution to the critical problem faced by many developing countries: predicting when and where food shortages will occur [10]. By identifying the most influential variables, it becomes possible to intervene proactively, reducing vulnerability and ultimately lowering the number of people affected by these crises.

Machine learning, a subfield of artificial intelligence, uses statistical methods to train models by learning from historical data. Through this process, algorithms derive parameter values that allow them to make more accurate predictions on unseen data. The better the quality of the training data, the more accurate the model's classifications will be.

Among the various models used in machine learning, one group is known as supervised learning, where the target variable is known, and the algorithm is given specific instructions about what to predict. Classification is one of the tasks that supervised learning performs. There are three main types of classification tasks: binary classification, multi-class classification, and multi-label classification.

1. **Binary classification** represents two possible outcomes, such as male or female, spam or not spam, or yes or no.
2. **Multi-class classification** involves more than two possible outcomes, which may be ordinal (ordered) or non-ordinal (without a natural order).
3. **Multi-label classification** allows for each instance to belong to multiple categories simultaneously.

The models are trained by feeding them with variables that represent the relationships between these inputs and the target variable. By understanding these relationships, machine learning models can help predict future outcomes with greater precision, aiding efforts to reduce food insecurity in vulnerable regions.

2.3. Food security predictions

In previous studies, several supervised machine learning algorithms—such as Support Vector Machines, k-Nearest Neighbours, Naïve Bayes, and Decision Trees—were tested for famine prediction. These models demonstrated accuracy rates of over 60% and ROC curves exceeding 90% [33].

In 2022, Deléglise et al. [12] employed heterogeneous data from various sources and formats—including surveys, GPS coordinates (such as hospitals, schools, and violent event locations), waterways, and raster images—to analyse two key indicators of food security: the Food Consumption Score (FCS) and the Household Dietary Diversity Score (HDDS). They proposed a framework for food security prediction based on heterogeneous data, using deep learning models to predict these two indicators .

Similarly, Modhurima Dey, Syed Badruddoza and Jill J. McCluskey [2] applied machine learning to predict the presence of healthful food retailers. In that study, they used machine learning to predict the modified Retail Food Environment Index, achieving a final accuracy of 72% within their sample data .

Predictions in food security are not limited to direct indicators, such as the Food Consumption Score or the number of people lacking three complete meals a day. They are also applied to related metrics that help identify key drivers of food shortages in specific regions. For instance, Lotfi et al. used machine learning as a framework for agricultural food capacity production forecasts. Their approach significantly reduced prediction errors compared to traditional models, demonstrating the importance of applied AI in today's-world sustainability [26].

Along the same lines, Harris Jabez applied neural networks in 2017 to forecast food prices in Canada [20]. His goal was to create a more accurate model for future Canadian food price reports and to monitor anomalies, resulting in a 1.6% improvement over previous models.

These examples demonstrate that machine learning predictions are widely used not only for food security indicators but also for adjacent topics that add substantial value in preventing food insecurity and reducing global hunger.

Identifying the combination of variables with the greatest predictive power is essential for enabling early interventions before food shortages arise. This helps reduce the number of people at risk and mitigates the potential domino effects triggered by food insecurity. Several studies have examined the effects and relationships between sociological, economic, and climatic factors on famine and food shortages. Although these relationships are complex and require extensive analysis, numerous studies highlight infrastructure [38], *armed conflict* [28], and climatic conditions [22] as direct contributors to food security. Additionally, factors such as *land degradation* [40] are critical to food production, a key component of food security. Changes in this metric can affect overall food production and food waste, leading to food insecurity in affected regions.

2.4. Granularity of the prediction, from local to global

The scale at which famine is analysed can present challenges when predicting indicators, as different AI techniques have been applied across various levels of granularity. These scales range from subnational [32], country [45], and multi-country [5] levels to continental and global scales. Although a global model may sacrifice some degree of accuracy and predictive power at the local level, its advantages include a comprehensive view of global trends, the discovery of potential patterns on a worldwide scale, and an inclusive methodology that accounts for multiple variables across different regions [4].

Despite the complexity, AI techniques have shown promising results across all scales, demonstrating positive outcomes and a clear path towards achieving the Sustainable Development Goal of Zero Hunger.

3. Methodology

3.1. Data

The classification proposed by the FAO's Global Information and Early Warning System on Food and Agriculture (GIEWS) defines three not mutually exclusive categories of food insecurity:

- Severe localised food insecurity
- Widespread lack of access
- Exceptional shortfall in aggregate food production/supplies

Farm-level data plays a crucial role in forecasting food security, as recognised by the Global Food Security Index (GFSI), which in 2022 incorporated 14 new metrics [4]. These metrics track the interconnection between food production and consumption. However, those measuring the relationship between food production and market accessibility, in terms of time and distance, are challenging to standardise and difficult to measure effectively.

The predictors (independent variables) were selected from various sources, as detailed in Table 2, which includes their respective source, units, and explanations. This selection was informed by the literature review. For instance, FAO data offers crucial insights into global food trends, covering areas such as production, pesticide use, soil types, crops, transportation, storage, and final consumption. The FAO's objective is to enhance food production and quality, reduce costs, mitigate hunger, and promote equality. One key FAO system, GIEWS, which monitors external food assistance, provided the target variable for this study. The target variable was binary—indicating whether or not a country required external food assistance in a given year.

The Economist Intelligence Unit (EIU) is another key resource, offering a wealth of economic data from 204 countries. This data, which is freely accessible, includes a range of metrics released annually, monthly, or even daily. For this study, two features, *road infrastructure* and *armed conflict* were included, following the recommendations of Jeremy Swift, who highlighted their importance in the food cycle. The infrastructure metric is presented as a category with five levels (0 to 4), where 0 represents the worst infrastructure, while the conflict metric similarly uses a scale where 0 indicates the lowest risk of *armed conflict*.
The World Bank Group, a unique global partnership committed to fighting poverty through sustainable solutions, provides diverse datasets covering banks, climate, food, economy, demography, development, energy, and employment, among others [47]. For this research, data on *forest change* and *Access to drinkable water* were used. The rationale is that healthy forests contribute to better water quality, which benefits agriculture, consumption, and overall quality of life [9].

Flooding data was obtained from ND-GAIN, an initiative of the University of Notre Dame that assesses where the greatest needs and opportunities for improving climate resilience exist [18]. It is well known that crops require specific climatic conditions and soil quality to thrive. High flood levels pose a significant risk to crops,

potentially leading to crop damage or destruction. Conversely, consistent average climate conditions reduce these risks.
The United Nations' data platform, UNData, provides a web-based service offering free access to a wide array of statistical databases. These cover various topics, including fertility, homicides, marriage, telecommunications, tourism, development, and *land degradation*. *Land degradation* refers to the reduction or loss of biological or economic productivity in arid, semi-arid, and dry sub-humid areas [17]. In this study, *land degradation* was assessed for its impact on the food cycle, as it directly affects the production phase and, consequently, plays a key role in food security and shortages.

Finally, the United Nations' annual report presents the widely recognised *Human Development Index*, a summary measure of average achievements in three key dimensions of human development: a long and healthy life, knowledge, and a decent standard of living. The *Human Development Index* is the geometric mean of normalised indices for these three dimensions, essentially providing a measure of quality of life by country or region in a given year.

Variable	Source	Units	Explanation
Risk	FAO	0 - 4 0: no risk 4: maximum risk	Countries requiring external assistance for food are expected to lack the resources to deal with reported critical problems of food insecurity.
Road infrastructure	EIU	0-4 0: worst 4: best	What is the quality of the national road infrastructure?
Forest change	World Bank	-100 - 100% <0: Less forest >0: more forest	A measure of the health of forests (change in forest areas as a percentage of total land area).
Flooding	Notre Dame Global Adaptation Initiative (ND-GAIN)	% change in flood hazard	Mean percentage change in flood hazard. Flood hazard is the monthly maximum precipitation in 5 consecutive days.
Food loss	FAO	1-100 % 1: lowest 100: highest	Total waste as a percentage of total domestic supply
Armed conflict	EIU	0 - 4 0: lower risk 4: highest risk	An assessment of the risk of armed conflict.
Land degradation	United Nations	Proportion of land that is degraded over total land area (%)	A measure of the proportion of land that is degraded over total land area
Human development index	Human Development Reports	1-100 % 1: lowest 100: highest	The Human Development Index is a summary measure of average achievement in key dimensions of human development
Access to drinking water	World Bank	1-100 % 1: lowest 100: highest	A measure of the percentage of people using safely managed drinking water services.
target	FAO	0 - 1 0: Does not 1: Does	Countries requiring external assistance for food
target_delayed	FAO	0 - 1 0: did not 1: did	Countries that required external assistance for food in the last year
Risk_delayed	FAO	0 - 3 0 : does not 3: Exceptional shortfall	Category in which each country that requires external assistance belongs to.

Table 2. The table defines and explains key variables used in the study, providing details on their names, sources, units, and explanation.

3.1.1. Data considerations

The selection of countries in this study does not reflect any socio-political or religious ideology. These countries were chosen solely based on the availability of data and their established boundaries. Therefore, the regions are presented for practical purposes of organising information and should not be interpreted as an endorsement of any particular ideological, political, or religious position.

3.2. Preprocessing

3.2.1. Missing values

The data used in this study came from various sources, each with its own format. Despite this, the dataset was generally complete. The primary dataset, imported from GIEWS, contained no missing values. It included details such as the date (month and year) when the alert was issued, the nature of the food insecurity, and the main causes behind the issue.

However, when concatenating the table with additional features, 2,702 missing values (NAs) were found (see Table 3). The breakdown of missing values for each variable is as follows:

Variable	**NAs**
country	0
year	0
Food loss	286
Road infrastructure	286
Forest change	286
Flooding	286
Armed conflict	286
Land degradation	286
Human development index	700
Access to drinking water	286

Table 3. This table presents the count of missing values (NAs) for each variable in the dataset after merging.

The first imputation was performed on the *Human Development Index*. Since all countries had *Human Development Index* values for the years 2012, 2015, 2019, 2020, 2021, and 2022, interpolation was chosen to impute the missing values for the years 2013, 2014, 2016, 2017, and 2018. As seen in the Human Development Report , the *Human Development Index* tends to follow a linear trend over time for each country [39]. Therefore, a linear interpolation method was used, filling in values between the known data points [46]. If this had not been the case, methods such as forward filling, backward filling, or model-based imputation would have been considered.

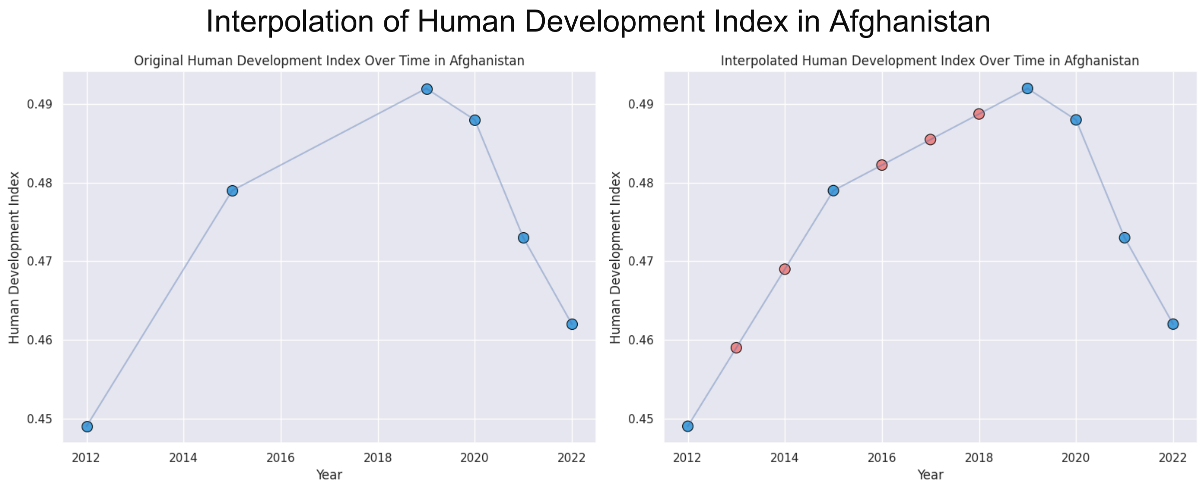

Figure 1. The figure shows the Human Development Index for Afghanistan from 2012 to 2022, with Figure 1 (left) displaying the original values, and Figure 1 (right) presenting the data after linear interpolation.

After this initial imputation, a custom function was created to fill the remaining missing values in the dataset. This function utilised the average of values from neighbouring rows, ranked according to the *Human Development Index* column. The approach works as follows:

1. **Sorting the Data**: The DataFrame is first sorted by country, year, and the specified ranking column ('*Human Development Index*').
2. **Iterating and Checking**: For each row, the function checks for missing values in columns other than the country, year, and *Human Development Index*.
3. **Imputing the Missing Values**:
    - If a missing value is found, the function looks at the previous and next rows for the same country.

- If both adjacent values are available, the missing value is imputed as the mean of the two.
- If only one neighbouring value is available, that value is used for imputation.

4. **Returning the Imputed Data**: The function then returns the DataFrame with missing values filled in.

For example, if Finland had no data on *food loss* for 2019, the function would interpolate the value using the average of Finland's data for 2018 and 2020. If Brazil lacked *food loss* data for both 2018 and 2019, the function would rank the countries by *Human Development Index* for those years, impute the missing data with values from countries that had similar *Human Development Index* scores, and repeat the process for 2019.

3.2.2. Outliers

Although most continuous variables in the dataset were expressed as percentages—such as *Forest change*, *Food loss*, and *Access to drinking water*—or as rankings from 1 to 100 like the *Human Development Index*, it was important to assess the presence of outliers across all variables. The *Human Development Index* was the only variable that did not present outliers, while the remaining variables exhibited some. However, the number of outliers in each variable was minimal and were managed as follows:

Variable	**Management**	**Number of outliers**
Forest change	Standardised so that the distribution was centred at 0 with a standard deviation of 1.	82
Food loss	Scaled using MinMaxScaler, normalising the range between 0 and 1.	45
Flooding	Not considered in the study (further explained in Chapter 3.3).	25
Access to drinking water	Scaled using MinMaxScaler to normalise the range between 0 and 1.	6
Human Development Index	Scaled so that the minimum was 0 and the maximum was 1.	0

Table 4: Here, the methods used to manage and scale the variables are shown, along with the number of outliers identified for each.

As shown in the preview table, the number of outliers was relatively low, with the highest percentage of outliers being less than 5% for *Forest change*. Although standardisation and scaling do not directly address the issue of outliers, they help mitigate the influence these values may have on the final model outputs by normalising the ranges of the variables (0-1) [42]. The overall distribution of the variables was left unchanged, primarily because of the low number of outliers. This decision allows for an analysis of how these outliers might impact model performance while maintaining the interpretability of the final models.

Another key reason for not altering the distribution of variables was that most of the machine learning models used in this study are not particularly sensitive to outliers. Models like decision trees, random forests, gradient boosting, and neural networks are robust to the presence of outliers. However, some models such as logistic regression, K-nearest neighbours (KNN), and neural networks (for training speed) do require scaling, hence the standardisation process for these variables.

3.2.3. Distributions

Some of the models employed in this study, such as Logistic Regression and Support Vector Machines (SVM), assume that the data follows a normal distribution [25]. Therefore, it was essential to evaluate the distribution shape of the variables used in the analysis.

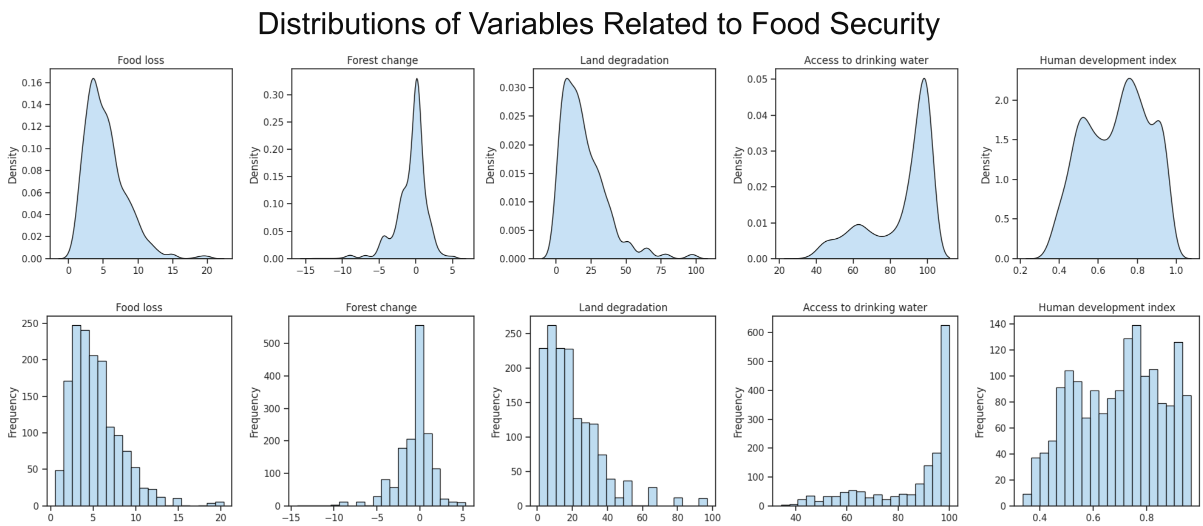

Figure 2. Density plots of continuous variables (top) and histograms (bottom) show the distribution of the variables: Food Loss, Forest Change, Land Degradation, Access to Drinking Water, and Human Development Index.

Upon examining the distribution plots, it was observed that the *Land degradation* and *Food loss* variables are slightly right-skewed. Similarly, the *Access to drinking water* variable is left-skewed, indicating that in most countries, the majority of the population has access to drinkable water.

While skewness in these variables does not pose significant issues for many of the models used, it could affect the performance of Logistic Regression and SVM, which are more sensitive to non-normal distributions [25]. To address this, a log transformation was applied to the right-skewed variables (*Land degradation* and *Food loss*) before running the models, while a root transformation was applied to the left-skewed *Access to drinking water* variable.

For the remaining models, such as decision trees, random forests, and neural networks, the data was passed into the models without transformation, as these models are generally robust to non-normal distributions.

3.2.4. Dealing with imbalance data

The dataset in this case exhibits a noticeable class imbalance, which can affect model performance if not properly addressed. To tackle this issue, three methods were tested: oversampling, undersampling, and SMOTE (Synthetic Minority Oversampling Technique) [11].

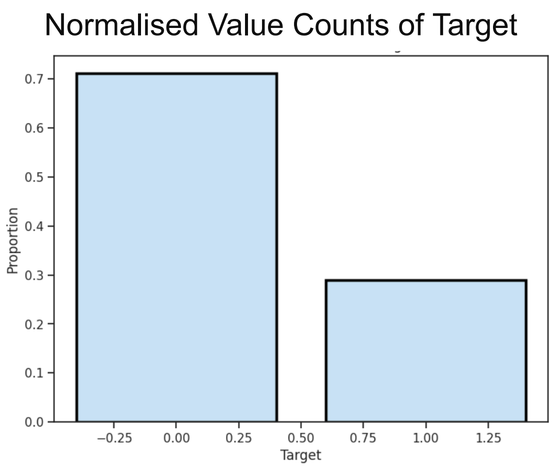

Figure 3. Bar plot of the normalised proportion of countries requiring external assistance for food (Target = 1) compared to those that do not (Target = 0). The categories are imbalanced, with approximately 70% of countries not requiring assistance, while 30% do.

- *Undersampling* involves randomly removing observations from the majority class until the dataset achieves balance. This method is particularly effective when the dataset is large, and even the minority class contains a substantial number of observations [11]. The drawback, however, is that it may eliminate important information by discarding data points, which could negatively impact the model's ability to learn all relevant patterns. Since the selection is random, valuable data might be lost during the process.
- *Oversampling* balances the dataset by replicating observations from the minority class until the two classes have an equal number of data points. While common, this technique risks overfitting as the same points are repeated multiple times. Additionally, it does not introduce any new information to the model, and the repeated data points may reduce variance, leading to suboptimal generalisation [11].
- *SMOTE (Synthetic Minority Oversampling Technique),* unlike traditional oversampling, generates synthetic examples of the minority class. SMOTE achieves this by selecting a data point from the minority class, identifying its k-nearest neighbours (set to 5 in this case), and generating a new, synthetic point based on a linear combination of the selected data point and its neighbours [11]. This method not only balances the dataset but also introduces new data that might enhance the model's understanding by adding diversity to the minority class.

A variant of SMOTE called BorderlineSMOTE was applied in this study. In this approach, the algorithm focuses on generating synthetic samples in the borderline area between the classes, where the likelihood of misclassification is higher. By targeting these areas, BorderlineSMOTE generates new points specifically where the minority and majority classes overlap or are likely to be confused [19]. It avoids generating points from noisy or irrelevant areas and creates synthetic data through linear interpolation between minority points in the borderline region and their k-nearest neighbours. This approach helps in training the model to better distinguish between the classes, especially near decision boundaries.

3.3. Feature selection

In the feature selection process, we aim to identify which variables contribute most to predicting the target variable, target_delayed, which represents whether a country required food assistance after a certain delay.

Variable	abs(correlation)
target_delayed	1.000000
Event	0.934353
target	0.931771
Risk	0.852528
Total events	0.796929
Access to drinking water	0.512978
Road infrastructure	0.371585
Armed conflict	0.334177
Food loss	0.207097
Forest change	0.139623
Flooding	0.067646
Human development index	0.052920
Land degradation	0.035577

Table 5. This table lists the absolute correlation values between the features and the ***target_delayed*** variable, showing the strength of the relationship each variable has with countries requiring external assistance in the previous year.

As shown in the table, several variables are highly correlated with target_delayed. The strongest correlation is with the Event variable, which is binary (0 or 1) and indicates whether a country needed food assistance in a given year. This is followed closely by Risk and Total events, reflecting the risk of needing assistance and the total number of assistance events a country has experienced.

On the other hand, *Land degradation* has the lowest correlation value, with an absolute correlation of approximately 0.04. Including this variable could introduce noise into the models, making the predictions less accurate. Additionally, its low correlation suggests it would not significantly improve the model's performance, while still making the model more computationally expensive. For these reasons, *Land degradation* was excluded from further analysis.

While *Flooding* is frequently cited in the literature as an important factor affecting food security, it shows a very low predictive power in this dataset [22]. Its correlation with other variables is minimal, indicating that it may not interact meaningfully with the other features. Including it in the models could introduce unnecessary noise. Thus, despite its theoretical relevance, Flooding was also excluded from the model.

Human Development Index, although showing a low correlation with the target variable, is retained in the analysis. *Human Development Index* is well-documented in the literature as a key indicator that captures a range of complex circumstances in a single summarised metric, which can help understand food security in significant ways [27]. While it may not be highly predictive on its own, it adds valuable context for interpreting the model's results. Moreover, the *Human Development Index* was transformed and scaled to enhance its predictive power. Tree-based models and neural networks are particularly adept at handling noisy or less informative variables, as they naturally perform feature selection during training [25]. Thus, the inclusion of *Human Development Index* does not negatively affect model performance, and its role in interpreting the results justifies its retention in the dataset.

3.4. Feature engineering

Variables are not always in the optimal format for model training, and in many cases, preprocessing can lead to the creation of new features that enhance model performance. Feature engineering is a crucial step where existing variables are transformed or combined to improve model accuracy and overall metrics. The most common techniques include addition, subtraction, modification, or deletion of features, all aimed at optimising predictive performance [14].

In this analysis, four new variables were created or modified using the initial features to extract more useful information and improve model outcomes:

- **Risk**: This variable was converted into an ordinal numeric variable to reflect a ranked ordering of food insecurity risk levels. The mapping is as follows:
    - 0 for no risk,
    - 1 for "Widespread lack of access",
    - 2 for "Severe localised food insecurity",
    - 3 for "Exceptional shortfall in aggregate food production/supplies."

- **Event**: This variable takes the value of 0 for all countries in all years unless a country required food assistance in a given year. In such cases, the variable records the number of times the country needed help in that year.
- **Total events**: This is the cumulative sum of the Event variable, representing the total number of food assistance occurrences a country has experienced up to that year, starting from 2009 (the year GIEWS began reporting).
- **Target_delayed**: This is the target variable to be predicted, which indicates whether a country will need external food assistance. It is created by shifting the original target variable by one year, so the model is trained to predict whether a country will require assistance in the following year based on current data.

These engineered features help to capture important trends over time, especially with Risk and Total events, and they allow the models to better account for the cumulative effects of food insecurity and related risks.

3.5. Modelling

This work addresses a classification problem using supervised models, specifically a binary classification task, where the goal is to predict whether a country will require external food supply (1) or not (0). Given this context, the choice of models is restricted to algorithms capable of handling binary classification, particularly those that have demonstrated promising results in similar studies. As many models as possible were explored in this work, taking into account time, complexity, resources, and available data.

The first model employed was Logistic Regression, a common algorithm in supervised learning that classifies the response variable into two categories. Instead of directly modelling a continuous Y variable, the model calculates the probability p that the target variable belongs to one of the categories (equation 1) [8]. The model's output is constrained between 0 and 1, where 1 represents the probability that the observation belongs to the default class—in this case, indicating that the country requires external food assistance. This consideration applies to all models used in this work. It is important to note that predictions depend on how the default class is initially set, making labelling a critical factor in interpreting the results.

The linear model attempts to represent a linear relationship between the predictors X and the response variable, as shown in the linear equation (equation 1):

$$p(X) = \beta_o + \beta_1 X \qquad \text{equation (1)}$$

Here, β_0 represents the intercept, or the baseline value of the response when the predictor X is zero, while β_1 represents the slope, which indicates how much the response variable changes for each unit increase in X. However, if a linear model were used in a binary classification problem, it could result in probabilities less than 0 and greater than 1 for very low and high values of X, respectively. To avoid this problem, Logistic Regression employs a method called maximum likelihood to fit the model, transforming the linear function and producing an output range constrained between 0 and 1 [8]. The logistic function used in Logistic Regression is calculated as follows:

$$p(X) = \frac{e^{\beta_o + \beta_1 X}}{1 + e^{\beta_o + \beta_1 X}} \qquad \text{equation (2)}$$

Generalising equation 2, we can predict a binary response using multiple predictors, with the equation rewritten as:

$$p(X) = \frac{e^{\beta_o + \beta_1 X_1 + \ldots + \beta_m X_m}}{1 + e^{\beta_o + \beta_1 X_1 + \ldots + \beta_m X_m}} \qquad \text{equation (3)}$$

Here, m represents the number of predictors included in the model, with each β_m representing the coefficient for its corresponding predictor X_m, converting equation 3 in the general form for m dimensional matrices.

Next, a Support Vector Classifier (SVC) was also employed. Support Vector Machines (SVM) are extensively used in machine learning after their discovery in 1990s and can be used for both classification and regression tasks. SVCs work by using hyperplanes that are regions of p-1 dimensions in a p-dimensional space to create boundaries for classification. Without involving deeply into the mathematical intricacies of the algorithm, as it is beyond the scope of this work, SVC uses hyperplanes to separate different classes. In binary classification, one side of the hyperplane corresponds to one class, while the other side corresponds to the other class. However, there are infinitely many possible hyperplanes, so the algorithm selects the maximal margin hyperplane, also known as the optimal separating hyperplane. This method finds the hyperplane that maximises the perpendicular distance from each training observation to the hyperplane. The region between the hyperplane and the closest data points is called the margin, and the algorithm seeks to find the largest margin that separates the classes [25].

K-Nearest Neighbours (KNN) is another algorithm employed in this study. It is a simple, non-parametric method used for classification in supervised machine learning tasks. Given a data point, KNN calculates the classes of the K nearest data as shown in equation 4 points to determine the probability that the observation belongs to a particular class, using the following equation:

$$Pr(Y = j|X = x_0) = \frac{1}{K}\sum_{i \in N_0} I(y_i = j) \qquad \text{equation (4)}$$

Where Pr is the conditional probability of class j as the fraction of points in N_0 whose response values are j. K is the number of neighbours considered, x_0 is the data point analysed and $I(y_i = j)$ is an indicator function that equals 1 if the i-th neighbor's response y_i matches class j, and 0 otherwise. The algorithm iteratively calculates this as can be seen in the summation of equation 4 until all observations are classified. The performance of this algorithm heavily depends on the distance metric used to calculate the closest data points, and it can become computationally expensive as it requires calculating distances between every data point iteratively [25].

Moreover, Decision Trees were used to classify the binary outcome using tree-based models, similar to other research in the context of famine prediction. In this non-parametric approach, a hierarchical tree structure consisting of nodes and leaves is used to classify each data point [25]. The nodes contain conditions that split the samples into binary groups until the algorithm can classify the majority of observations correctly. The function that measures the quality of a split is a crucial hyperparameter in a decision tree classifier, along with the splitter that determines where a node will be created, which can be random or deterministic. One of the most common splitters is the Gini impurity, calculated as shown in equation 5. This is not the only way to evaluate the nodes, however it is the most common approach.

$$gini\ impurity \ = \ 1 - \sum_{i=1}^{j} (p(i) \cdot (1 - p(i))) \qquad \text{equation (5)}$$

Where $p(i)$ is the probability of a specific class and j is the number of classes. The alternative with the lower Gini impurity is considered the best split for the model.

This intuitive yet powerful model can become more complex when a maximum depth parameter is introduced, which restricts the number of nodes a branch can have. Too many nodes can lead to overfitting, where the model performs well on training data but poorly on unseen data. Conversely, too few nodes can lead to underfitting, where the model fails to capture the training data patterns adequately. Some advantages of decision trees are that they are relatively easy to understand, implement, and interpret, as they are built using a series of if-statements. This simplicity is a significant advantage when it comes to model evaluation and result interpretation.

Another tree-based model used was Random Forest, which combines multiple decision trees to produce a single, more robust tree. While decision trees often perform well on training data, they tend to overfit and perform poorly on unseen data. Random Forest mitigates this by creating several decision trees and using a criterion called majority voting to form the final decision. This approach improves accuracy and generalisation on unseen data [25]. In Random Forest, when a new observation is classified, the algorithm passes the data through each tree, and the majority vote determines the final classification. The process of bootstrapping and aggregating trees to make a decision is known as Bagging.

Gradient Boosting, another ensemble technique, was also employed. Unlike other methods that learn from data independently, Gradient Boosting combines several weak models to create a stronger, more sophisticated model with better performance. To select the best model, a loss function is calculated for each one. In machine learning, a loss function quantifies the difference between the model's predictions and the actual values [44]. A lower loss function indicates a better model, as the distance between the predicted and actual values is minimal. In binary classification, the loss function used is typically the cross-entropy function, which measures the difference between two probability distributions rather than the values themselves.

Finally, Neural Networks were employed. These are a type of architecture that takes a vector as input, with the same length as the number of predictors, and passes it through hidden layers containing units called neurons. These neurons convert linear functions into more complex nonlinear functions with the help of activation functions (equation 6 and 7), enabling the model to be non-linear. Most neural networks follow the general form:

$$f(x) = \beta_0 + \sum_{k=1}^{K} \beta_k h_k(X) \qquad \text{equation (6)}$$

Where:

$$h_k(X) = g(w_{k0} + \sum_{j=1}^{p} w_{kj} X_j) \qquad \text{equation (7)}$$

Here, $f(x)$ represents the output of the neural network, β_0 is the bias term, and β_k are the weights for each neuron in the hidden layer. The function $h_k(X)$ represents the output of the k-th hidden neuron, where g is the activation function, w_{k0} is the bias term for the k-th neuron, and w_{kj} are the weights applied to each predictor X_j.

Neural networks have achieved tremendous success in various fields such as computer vision, image recognition, and generative AI. However, they are challenging to train, and finding the optimal parameters can be a time-consuming and resource-intensive task. Unlike traditional machine learning techniques, neural networks require tuning many more hyperparameters, such as learning rate, number of layers, number of neurons per layer, activation functions, batch size, dropouts, Batch Normalisation, epochs, and patience in early stopping. All these factors make neural networks computationally expensive and limit their use in terms of time and resources. Moreover, while neural networks perform well, they require large amounts of data for training and are not suitable for every task, especially simpler tasks where they may overfit by learning patterns too well, failing to generalise to unseen data.

3.6. Hyperparameter tuning

When tuning a model, it is crucial to consider which hyperparameters can be adjusted and to account for the available time and computational resources. Among the most common techniques for hyperparameter tuning are Manual Search, Bayesian Optimisation, genetic algorithms, Grid Search, and Random Search. The latter two are the most widely used in industry due to their relative ease of implementation. The primary difference between Random Search and Grid Search lies in the search space used to find the optimal parameters for the model, which directly affects computational power requirements. Random Search has a fixed number of combinations to be tried, which can be set by the user according to the resources and time available. In contrast, Grid Search exhaustively explores the entire parameter space [34].

Although Random Search may involve fewer searches than Grid Search, both theoretical evidence (from probability theory) and empirical results from numerous trials have demonstrated that the chances of achieving comparable model performance are very high as the number of random combinations increases. When time and computational power are considered, the trade-off is significantly in favour of Random Search. Therefore, Random Search was employed in this project for all the models. The hyperparameters used are shown in the following table, along with the number of Random Searches carried out. The final parameters are presented in the Results section.

Algorithm	Hyperparameters	Total iteration
Logistic regression	'penalty' : ['l1', 'l2'] 'C' : np.logspace(-4, 4, 20) 'solver' : ['liblinear']	40
Suported vector machines	kernel: ["linear", "poly", "rbf", "sigmoid"] "C": [0.1, 1, 10] "gamma": [0.00001, 0.0001, 0.001, 0.01, 0.1]	40
K-neares neighbours	"n_neighbors": np.linspace(1, 30, 30).astype(int), "algorithm" : ['auto', 'ball_tree', 'kd_tree', 'brute'], "leaf_size": np.linspace(1, 50, 6).astype(int), "p": [1,2]	200
Decision trees	'criterion' : ['gini', 'entropy'], 'splitter' : ['best', 'random'], 'max_depth' : [1, 2, 3, 4, 5, 6,7,8,9,10], 'min_samples_split' : [2, 3, 4, 5, 6, 7, 8, 9, 10]	300
Random Forest	'bootstrap': [True, False], 'max_depth': range(2,20,2), 'max_features': ['log2', 'sqrt'], 'min_samples_leaf': [1, 2, 4], 'min_samples_split': [2, 5, 10], 'criterion' : ['gini', 'entropy'], 'n_estimators': [50,100, 200, 300]	250
Gradient boosting	'learning_rate': [0.01, 0.05, 0.1, 0.2], 'max_depth': [3, 4, 5, 6], 'subsample': [0.6, 0.7, 0.8], 'colsample_bytree': [0.6, 0.8, 1.0], 'gamma': [0, 0.1, 0.2], 'reg_alpha': [0, 0.1, 1], 'reg_lambda': [1, 10]	200
Multi-layer Perceptron	'neurons_imp_layer': [from 32 to 512] 'activation_input': ["relu", "tanh"] '1st_dropout': [True, False] 'n_hidden_layers': [1,2] 'neurons_1_layer': [from 32 to 512] 'neurons_0_layer': [from 32 to 512] 'activation_0_hidden': ["relu", "tanh"] 'BatchNormalization': [True, False] '2nd_dropout': [True, False] 'learning_rate': from 1e-5 to 1e-2 'activation_1_hidden': ["relu", "tanh"] 'Dropout': from 0.1 to 0.2	50

Table 6. This table outlines the hyperparameters and total number of iterations used for tuning various machine learning algorithms in the study.

3.7. Model evaluation and metric selection

In many cases, particularly in fields where identifying positive cases is critical—such as medicine, biology, and pharmacology—recall is often prioritised over accuracy as the key metric for tuning and selecting the best models. The rationale is that the consequences of misclassifying positive cases can be severe, even fatal (see equation 8). There is no universal guideline for this; it always depends on the specific context, the purpose for which the algorithm is developed, and the goals that the scientist has in mind for the model.

Recall is defined as:

$$Recall = \frac{TP}{TP+FN} \qquad \text{equation (8)}$$

Where:

- **TP (True Positive)** refers to countries that require external assistance for food and were correctly classified by the model.
- **FN (False Negative)** refers to countries that require external assistance for food but were incorrectly classified as not needing assistance by the model.

In this study, recall (or sensitivity) is prioritised over accuracy as was done in other studies like the one of Irshad et al. [23] because the focus is on analysing the true positive rate. Consider a scenario where a model predicts that Country A will experience a food shortage next year, even though, in reality, it will not. As a result, the country might take preemptive measures, closely monitor processes within the food supply chain, and pay particular attention to areas with higher vulnerability. The outcome would be a country that avoids a food shortage and a more robust food supply system, reducing the risk of future crises.

On the other hand, imagine the opposite scenario, where the model predicts that Country A will not face a food shortage, but in reality, the country does experience one. In this case, the country might become complacent, redirecting resources and efforts to other areas, thereby neglecting the necessary precautions in the food supply chain. This scenario is precisely what must be avoided, as it could lead to catastrophic outcomes, potentially putting thousands or even millions of lives at risk due to a misinterpretation of the model's metrics.

4. Results and analysis

4.1. EDA

After the preprocessing phase, the dataset comprised 2,162 observations, with an equal distribution between the two classes. This is a relatively large dataset, especially considering that the predictions are made yearly and span 139 countries over 12 years.

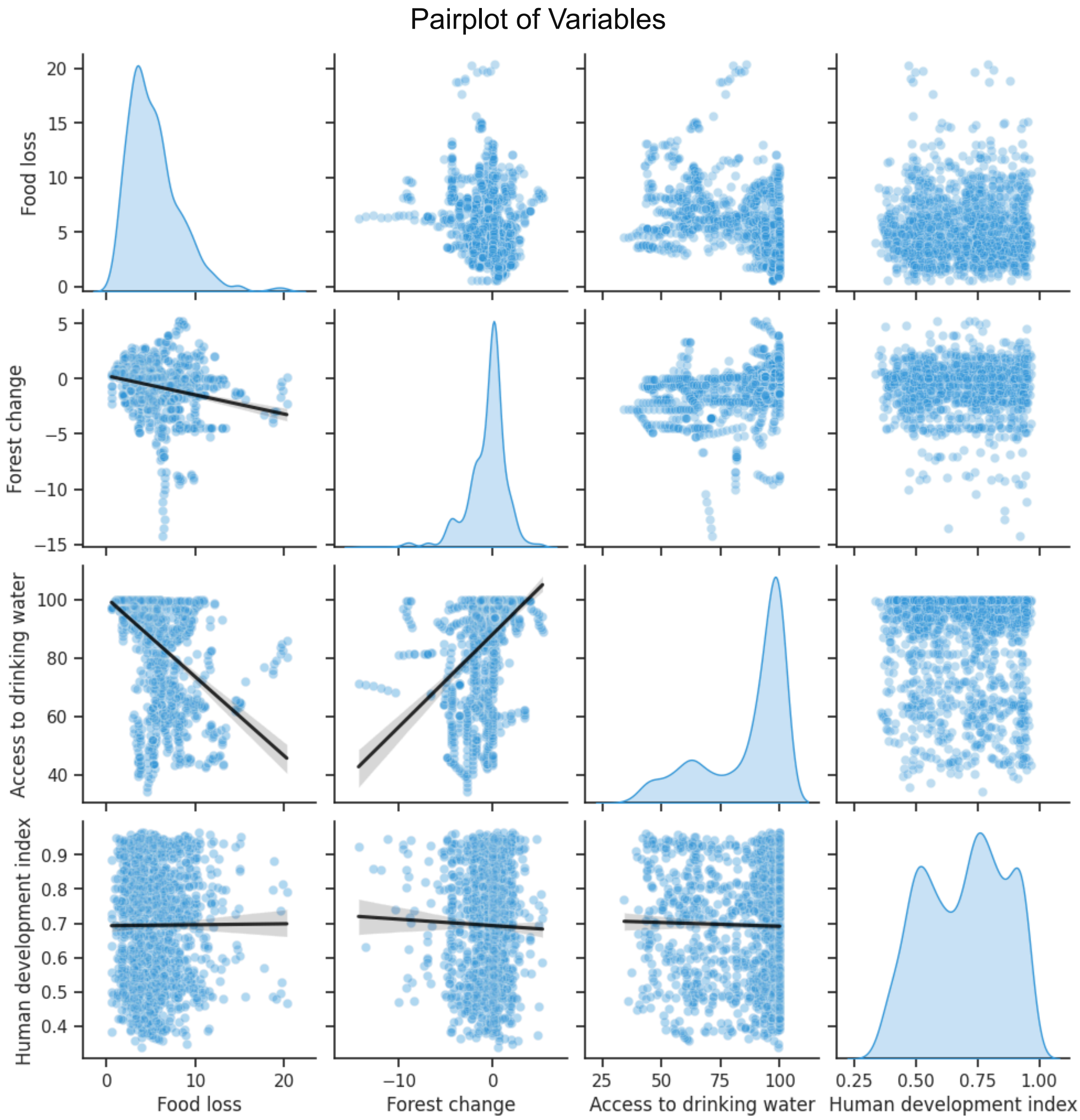

Figure 4. This pairplot visualises the relationships between continuous variables: Food Loss, Forest Change, Access to Drinking Water, and Human Development Index. The diagonal displays the KDE plots for each variable, while the scatterplots in the lower triangle show pairwise comparisons with linear regression lines.

Initially, it was observed that the *Food Loss* variable exhibited a slightly right-skewed distribution. This suggests that in most countries, *food loss* remains relatively low for the majority of the years. This observation aligns with the work of Ishangulyyev, Kim and Lee in 2019 where they analyse the problem of the food chain and conclude that despite the large quantity of food being wasted, the majority of all food produced is still consumed and one third of it is wasted [24]. While it would be interesting to investigate whether *food loss* has been increasing or decreasing over the years, such analysis falls outside the scope of this study. It is important to note that the *Food Loss* axis ranges from 0 to approximately 25, indicating that *food loss* has never exceeded a quarter of total food production over the last 12 years globally.

The relationship between *Food Loss* and variables such as *Forest Change* and *Access to Drinking Water* is also interesting. In both cases, as *food loss* decreases, these two variables also decrease. The relationship with *Access to Drinking Water* is intuitive. Less access to water implies that food cannot be washed or stored properly, leading to increased *food loss*. However, the relationship with *Forest Change* is less straightforward. As explained by D. Ellison and other researchers in 2017, Forests can be used for many ecosystem services, one of them is the mitigation of problems related to water scarcity and global warming [15]. In many cases, the negative value of *Forest Change* is due to deforestation for agricultural purposes, which should theoretically increase food production and improve food security at the expense of the local ecosystem.

The *Forest Change* variable follows a tight normal distribution, slightly skewed to the left, but still treatable as a normal distribution. The median is -0.01, and the mean is -0.64. This discrepancy is due to the presence of outliers. While the general trend suggests that forest areas have remained largely unchanged, even a small change in this variable could translate into millions of hectares affected, leading to potentially devastating local and global effects and affectation not only in energy but also in water flow as exposed in 2016 for R Alkama and A Cescatti [1]. Regarding Access to Water, the behaviour is intuitive: the more natural forest cover, the greater the access to water, as the majority of the water cycle is sustained by native forests and mountainous regions.

Access to Drinking Water is left-skewed, with a median of 94 and a mean of 86.12, suggesting that, on average, most countries have had high levels of *Access to drinking water* over the last 12 years. However, the thick tails of the distribution highlight significant inequality, which aligns with the findings from the 2019 research by H. Ritchie, F. Spooner and M. Roser [36]. Their work demonstrated that unsafe water remains one of the largest health and environmental problems, particularly affecting the poorest populations. Lack of access to safe water is a leading risk factor for infectious diseases such as cholera, diarrhoea, dysentery, hepatitis A, typhoid, and polio, and it exacerbates malnutrition and childhood stunting. The Global Burden of Disease study, published in The Lancet [31], ranks unsafe water as a significant risk factor for death worldwide. The distribution ranges from 25% to 100% of the population, indicating the highest variance among the three previously analysed variables and revealing a significant gap in access to essential resources among countries. This variance is crucial for the model, as the variable could provide valuable information for prediction, with a correlation of -0.5 suggesting that it could effectively differentiate between countries with varying levels of food security.

Lastly, the *Human Development Index* exhibits a multimodal distribution, with peaks around 0.5, 0.75, and 0.9. As seen in the pairplot and summarised in the accompanying table, the relationship between *Human Development Index* and other variables is weak. However, it is important to notice that *Human Development Index* has the highest variance of all the numeric variables, which could be particularly useful in cases where the models struggle to classify, especially in borderline cases where the samples are similar and prone to misclassification.

Regarding data imbalance, all models in this study were evaluated using four techniques: Oversampling, Undersampling, SMOTE, and Borderline SMOTE. It is important to note that no single technique is superior to another; the best approach depends on the characteristics of the dataset and the models being used. More complex techniques are not always better, as performance gains depend on the specific nature of the data. This can be demonstrated by comparing the results of this study with similar works in the field. For instance, unlike the approach taken by AM Soleh and B Susetyo (2024) [37], where they addressed class imbalance using SMOTE and other algorithms to predict the food insecurity status of households, in this study, SMOTE outperformed both oversampling and undersampling techniques.

In real-world food security predictions, it is common to encounter imbalanced datasets where the number of countries requiring assistance is much greater than those that do not. This creates challenges when adjusting the sample size by either increasing the minority class (countries not needing assistance) or decreasing the majority class (countries needing assistance).

Reducing the majority class through random undersampling can cause the loss of valuable information, making it difficult for models to capture general patterns in the data. In machine learning, larger datasets tend to improve predictive performance, provided the data quality is maintained. In this study, random undersampling resulted in poor model performance during both training and testing, suggesting that the dataset became insufficient for the models to generalise effectively.

On the other hand, oversampling the minority class yielded better accuracy than undersampling. However, metrics like sensitivity and specificity declined, and the models struggled to identify true positives. This decline in performance is expected since oversampling repeatedly resamples the predictors for the minority class, leading to a reduction in data variance. As a result, the models became more adept at classifying the false class while underperforming when trying to correctly identify true positive instances.

In this case, the Borderline SMOTE technique was proven to be the most effective resampling method. While the overall accuracy was comparable to that achieved with oversampling, recall and sensitivity were significantly higher, indicating that the models performed better at identifying true positives. This suggests that focusing on the borderline regions where misclassification is more likely provides a more effective strategy for handling imbalanced data in this context. Borderline SMOTE's ability to generate synthetic samples near the decision boundary ensures that the models better capture the complexity of the minority class, leading to improved performance in many unbalanced scenarios [19].

4.2. Model performance

As outlined in the preprocessing chapter, the model performance was primarily evaluated using the recall metric, which is crucial for this study's context due to the need to accurately identify countries that require external assistance for food security. In this domain, missing a country that needs help would have far more severe consequences than incorrectly identifying a country that does not. Recall, therefore, is the priority to ensure the model maximises true positives (i.e., correctly identifies countries needing assistance).

Model	Recall	Accuracy	Precision	ROC_AUC
Gradient Boosting	0.978788	0.975347	0.972892	0.975287
Random Forest	0.975758	0.976888	0.978723	0.976907
Decision Trees	0.975758	0.972265	0.96988	0.972205
k-Nearest Neighbours	0.975758	0.959938	0.947059	0.959666
Multi-layer Perceptron	0.966667	0.969183	0.972561	0.969227
Logistic Regression	0.954545	0.898305	0.860656	0.897335
Support Vector Machines	0.918182	0.946071	0.974277	0.946552

Table 7. It summarises the performance of various machine learning models based on key evaluation metrics: Recall, Accuracy, Precision, and ROC_AUC.

As observed in table 7, Gradient Boosting model emerged as the top performer in terms of recall, achieving a score of 0.978, meaning it successfully identified almost all of the countries that required assistance. This high recall value is accompanied by a high accuracy (0.975) and precision (0.973), which highlights the model's robustness not only in identifying the correct countries but also in minimising the false positives. The model's high ROC_AUC score (0.975) demonstrates its ability to distinguish between the classes (countries needing assistance vs. those not needing assistance) across various decision thresholds, making it adaptable for different operational requirements.

The Random Forest model, with a recall of 0.975, shows nearly identical performance to Gradient Boosting in terms of recall but a slightly higher accuracy (0.976) and precision (0.979). The Random Forest's strength lies in its ability to minimise errors across multiple evaluation metrics, which makes it a highly reliable option for predicting countries at risk of food insecurity. The random sampling and combination of decision trees in this model help it capture complex patterns in the data, likely reflecting interactions between features like Risk, Access to drinking water, and *Armed conflict*.

Decision Trees, although slightly less accurate than the ensemble methods, still achieved a recall of 0.975, meaning it performed well at identifying the high-risk countries. However, its slightly lower precision (0.970) and accuracy (0.972) suggest that the standalone decision tree model might be more prone to misclassifying countries that don't need assistance, which could lead to resource misallocation.
k-Nearest Neighbors (kNN) achieved the same recall as Random Forest and Decision Trees (0.975), but with a notably lower accuracy (0.960) and precision (0.947). This discrepancy highlights kNN's vulnerability to misclassifying some countries, likely due to its sensitivity to local patterns in the data rather than global trends. Since food security data is likely highly variable across regions, kNN might struggle to generalise effectively.
Logistic Regression and Support Vector Machines (SVM) were the least effective models in this context, with both recall and accuracy lagging behind the more complex models. Logistic Regression, with a recall of 0.955 and an accuracy of 0.898, suggests that this linear model struggles to capture the complex interactions between variables like Risk, Road infrastructure, and Land degradation. Similarly, SVM's recall of 0.918, despite a relatively high precision (0.974), indicates that it is less suitable for maximising true positive identification. These models are likely unable to handle the non-linear and high-dimensional aspects of food security data.

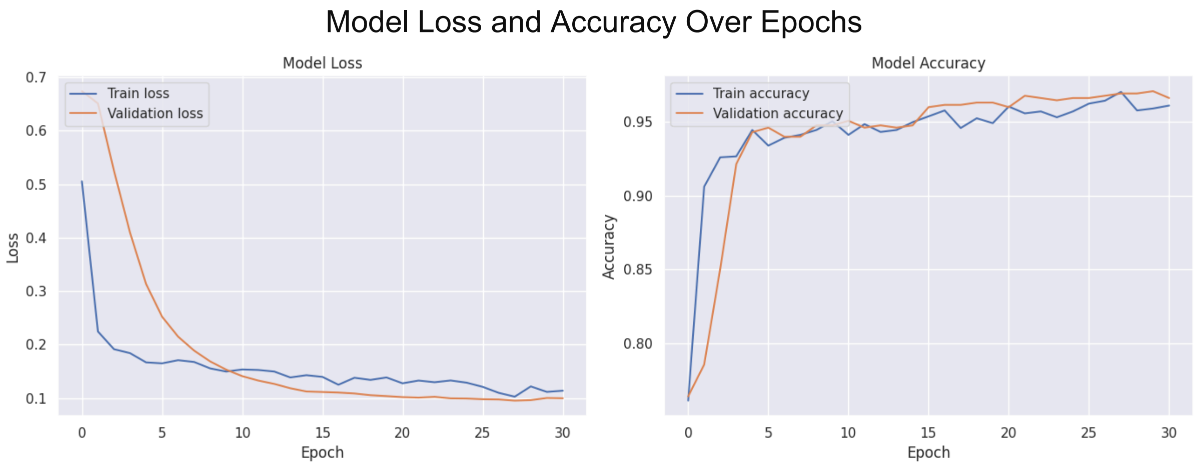

Figure 5. The figure shows the MLP model's performance over 30 epochs, with both training and validation loss (left). Similarly, training and validation accuracy are shown over the training epochs (right).

The Multi-layer Perceptron (MLP) performed slightly worse in terms of recall (0.967) compared to the ensemble methods but maintained a respectable accuracy (0.969). The model's capacity to capture non-linear relationships in the data is beneficial in cases where interactions between variables like Forest change, *Food loss*, and *Human Development Index* are complex. However, the slight reduction in recall suggests it might miss identifying a small subset of countries that need assistance.

As shown in the training and validation curves, the model's loss decreased consistently over the epochs, both for the training and validation datasets, indicating that the model was learning effectively. The train loss and validation loss converge smoothly, showing that the model did not overfit, a common challenge for deep learning models like MLPs. By epoch 10, both losses stabilise, which implies the model reaches an optimal point after sufficient learning.

The accuracy curves also support this observation. The training accuracy quickly climbs, followed closely by the validation accuracy, both reaching values near 0.95. The close alignment between training and validation accuracies further suggests the model generalises well to unseen data, with minimal overfitting. The slight fluctuations in validation accuracy after epoch 15 are relatively minor and show that the model maintains strong performance throughout the training process.

4.3. Model comparison

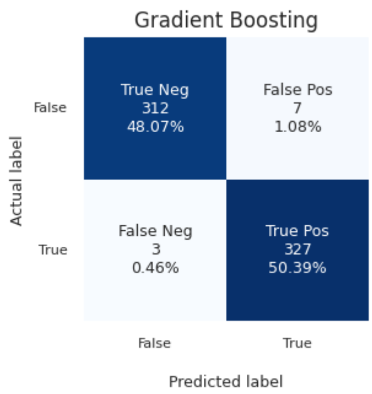

Figure 6. Confusion matrix showing the performance of the Gradient Boosting model.

When comparing the models based on their performance across various metrics, Gradient Boosting and Random Forest stand out as the most effective models for predicting food insecurity. Both models demonstrate high recall and accuracy, which are critical for this study's goal of identifying countries that require external assistance. These ensemble methods excel at capturing complex relationships in the data through their ability to aggregate the results of multiple decision trees, allowing them to generalise well across diverse datasets. The high ROC_AUC scores for both models (0.975 for Gradient Boosting and 0.977 for Random Forest) reflect their robustness across different classification thresholds, making them versatile options for operational decision-making. The slight differences between Gradient Boosting and Random Forest lie in their precision and accuracy. Random Forest shows marginally better accuracy (0.977) and precision (0.979) compared to Gradient Boosting (0.975 and 0.973, respectively). This implies that while both models are effective at identifying true positives, Random Forest may be slightly better at minimising false positives. Given the importance of resource allocation in food security, where incorrectly identifying a country as needing assistance can lead to wasted resources, this higher precision may make Random Forest a slightly more reliable choice in scenarios where the cost of false positives is high.

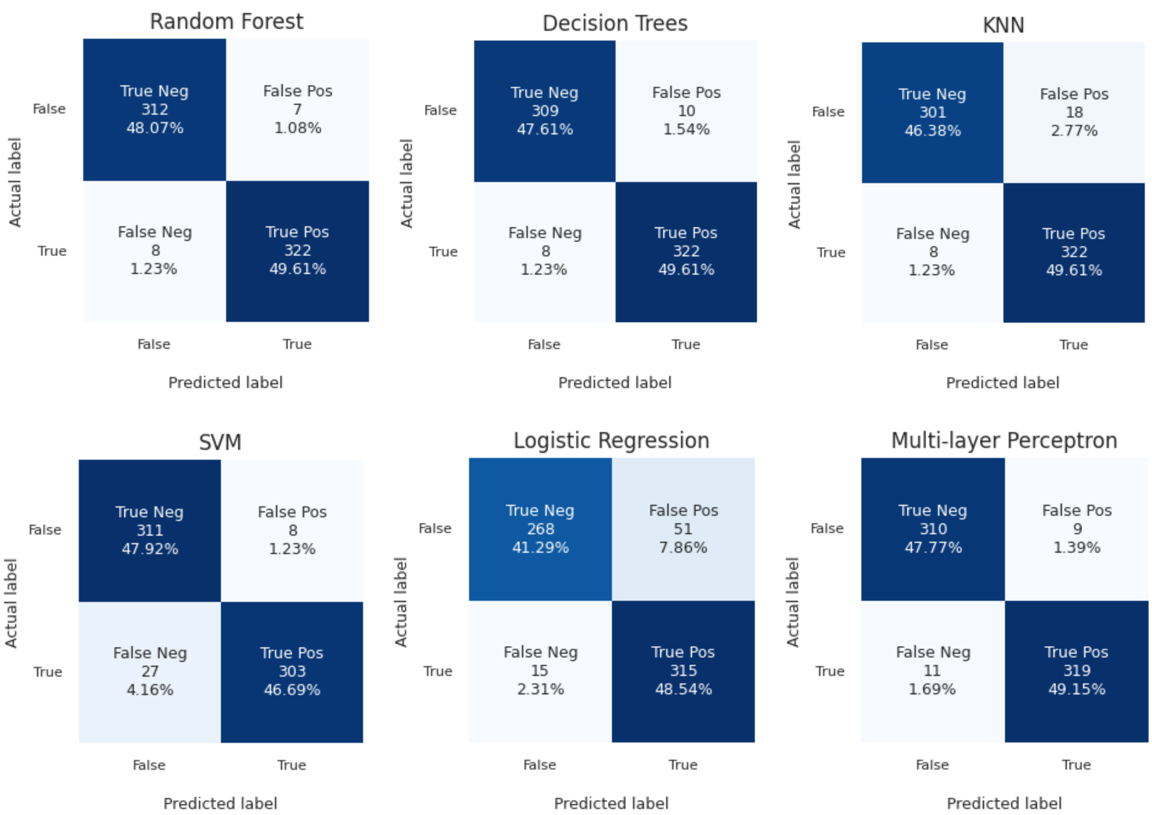

Figure 7. This figure presents the confusion matrices for six machine learning models: Random Forest, Decision Trees, k-Nearest Neighbors (kNN), Support Vector Machines (SVM), Logistic Regression, and Multi-layer Perceptron (MLP). Each matrix illustrates the balance between true and false classifications for countries requiring and not requiring external assistance

This is also supported by the confusion matrix for Random Forest , where the model has a very low number of false positives (7) and false negatives (8), leading to a total recall of 0.976 and a strong balance between identifying true positives and minimising incorrect classifications (Figure 7). With 312 true negatives and 322 true positives, Random Forest offers high classification accuracy while maintaining low misclassification rates.

Decision Trees also perform well in terms of recall (0.976), indicating that it can effectively identify the countries most at risk. However, its slightly lower accuracy and precision suggest that it may be more prone to overfitting or making classification errors compared to the ensemble methods. The confusion matrix for Decision Trees shows 10 false positives and 8 false negatives, which is slightly higher than the Random Forest model. This reflects the slightly higher likelihood of making errors, especially in predicting countries that do not need assistance (leading to false positives). However, Decision Trees remain valuable due to their simplicity and interpretability, allowing policymakers to understand the decision-making process.

k-Nearest Neighbors (kNN), despite achieving the same recall as Random Forest and Decision Trees (0.976), performs less favourably in terms of accuracy (0.960) and precision (0.947). The confusion matrix for kNN reveals a higher number of false positives (18) compared to Random Forest and Decision Trees, indicating a greater likelihood of incorrectly classifying countries as needing assistance. This higher false positive rate, reflected in a precision of 0.947, suggests that kNN is more prone to misclassifying countries that don’t require food aid. Misallocating resources based on these false positives could divert aid from countries in greater need, making kNN less reliable in this specific context.

The Multi-layer Perceptron (MLP) model strikes a balance between recall (0.967) and accuracy (0.969). Its strength lies in its ability to model non-linear interactions between variables like *Land degradation*, *Forest change,* and *Human Development Index*, which may not be captured as effectively by simpler models. However, the slight drop in recall compared to the ensemble methods suggests that MLP may miss identifying a subset of at-risk countries. The complexity of the model also makes it less interpretable than Decision Trees, though it may be more suitable in cases where the data relationships are too intricate for simpler models to capture.

Logistic Regression and Support Vector Machines (SVM), despite being widely used for classification tasks, fall short in this context. Logistic Regression's recall of 0.955 and accuracy of 0.898 highlight its limitations in capturing complex, non-linear relationships. The low accuracy suggests that Logistic Regression misclassifies a significant number of countries, which could be due to its inability to handle the high-dimensional nature of the food security data. Similarly, SVM, while maintaining a high precision of 0.974, has a recall of just 0.918, meaning it fails to identify many countries that need assistance. SVM's reliance on finding a hyperplane to separate the classes may not be as effective when the data contains overlapping or non-linear patterns, as seen in food security datasets.

4.4. SHAP

As explained by Christoph Molnar (2020) [30] Shapley values are the average marginal contribution of a feature value across all possible coalitions, so that provides important insights into how individual features contribute to a given ML model's predictions, helping to interpret complex models by quantifying the impact of each feature on the outcome.

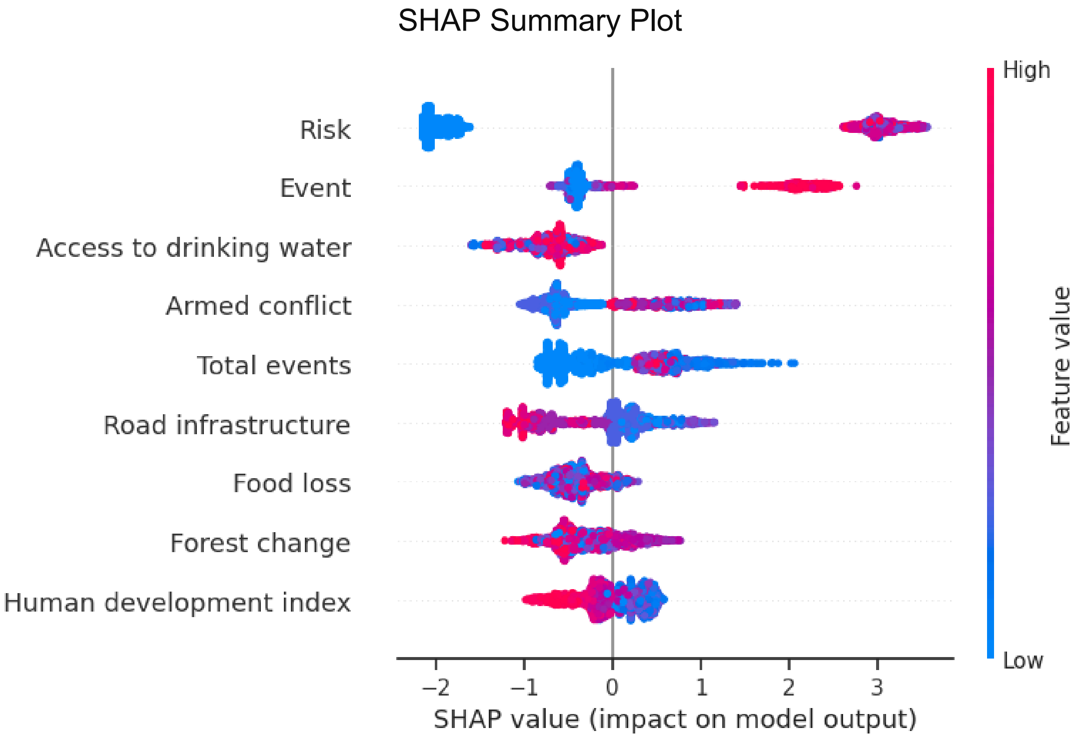

Figure 8. SHAP summary plot illustrates the impact of key features on the model's output for predicting countries needing external food assistance. Each dot represents a SHAP value for an individual prediction, with colours indicating the feature values (blue for low, red for high). The horizontal position shows the effect of the feature on the model output, with values further from zero having a larger influence.

The SHAP Summary Plot (figure 8) offers a distribution of how each feature contributes to the model's prediction across all countries in the dataset. The horizontal axis represents the distribution of the feature's contribution to the model output. This plot shows not only the magnitude of the impact but also the direction (positive or negative). For example, Risk shows a clear and consistent positive contribution, meaning that countries with higher risk scores are significantly more likely to face food insecurity. The density of red points (indicating high feature values) concentrated on the right side of the plot underscores how higher risk is strongly associated with increased food insecurity. In contrast, *Access to drinking water* demonstrates a broader spread across both positive and negative SHAP values, suggesting that access to potable water has a more complex relationship with food insecurity. The blue dots (lower access to water) are largely aligned with positive SHAP values, reinforcing that limited access increases the risk of food insecurity, while better water access (red dots) tends to mitigate risk.

Another interpretation of the variable *Access to drinking water* reveals the context-specific nature of its influence. While the majority of instances with poor water access correspond to higher food insecurity risk, which has been seen in different works and appears to be key not only in the models built here but also analysis in the field [13], the presence of a few red points with positive SHAP values indicates that in some countries, even where water access is relatively higher, other interacting factors may still push food insecurity risk upward. This could be due to interacting variables like infrastructure, *food loss*, or political instability or conflict, which reduce the mitigating effect of improved water access.
The *Event* feature (figure8), similarly, shows a bimodal distribution. Some countries experience a strong positive impact from this variable, indicating that certain types of events (such as economic downturns, political instability, or natural disasters) are key drivers of food insecurity. However, a few blue points on the left (negative SHAP values) suggest that in some cases, events may reduce food insecurity risk, possibly indicating instances where proactive responses or international aid successfully mitigated the negative impact of these events. This variability highlights the importance of understanding the type and scale of events and the resilience or capacity of countries to respond effectively.

SHAP waterfall plot

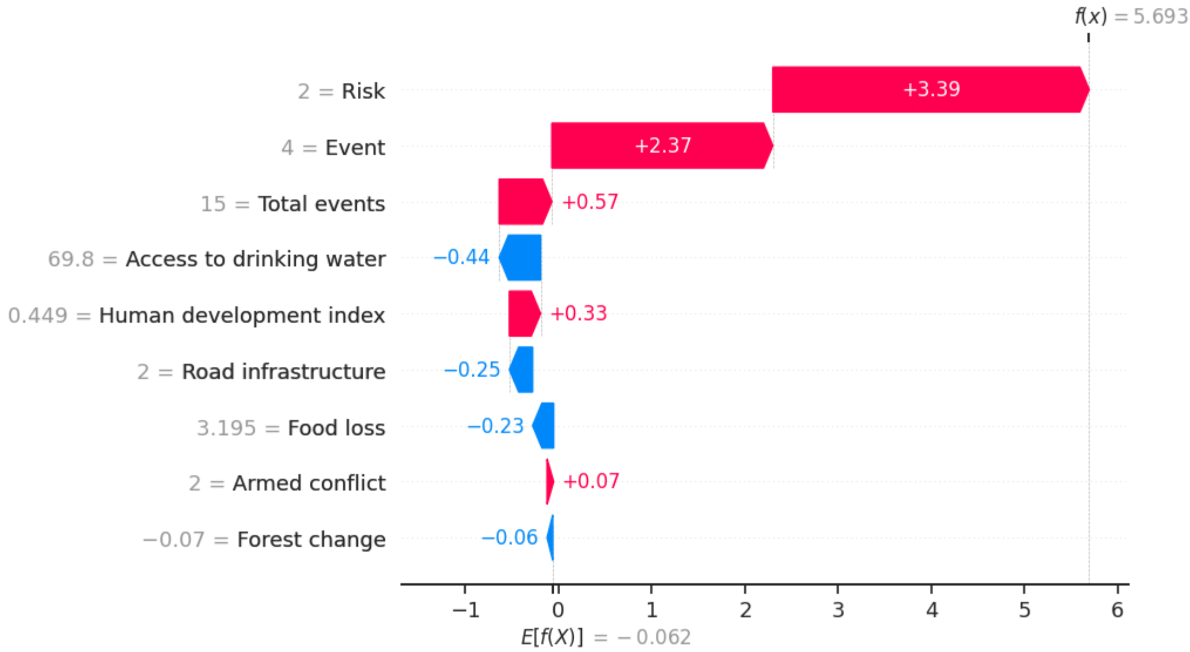

Figure 9. Illustrates how different features contribute to the predictions of the model, starting from the model's base value. Each bar represents a feature's impact on the model output, pushing the prediction higher (positive contribution, in red) or lower (negative contribution, in blue).

The SHAP Waterfall Plot (figure 9) provides a view of how each feature cumulatively adjusts the model's base prediction for a specific country. The base value, which represents the average food insecurity risk across all countries, is adjusted up or down by the contribution of each feature. In this case, *Risk* is the dominant positive contributor, adding over 3 SHAP points to the prediction. This feature's influence suggests that the country is facing extreme systemic risks, whether they be environmental, economic, or social, that make food insecurity highly likely. Event adds a smaller but still significant positive contribution, indicating that this country has experienced one or more events that exacerbate food insecurity. Together, these two features drive the prediction upward, signalling that the country is at high risk.

Interestingly, *Access to drinking water* and *Human development index* have a negative impact on the model's prediction, decreasing the overall food insecurity risk. The negative SHAP value for *Access to drinking water* suggests that in these countries, access to clean water is relatively high, which helps to buffer against food insecurity. *Human development index*, which encapsulates factors like education, income, and health, similarly reduces the overall risk, indicating that the country's development level is providing some protection against food insecurity. However, the relatively smaller magnitude of these negative SHAP values compared to the positive contributions from Risk and Event suggests that while development and water access help, they are not sufficient to fully mitigate the country's overall vulnerability to food insecurity. This interplay between features, with some acting as risk amplifiers and others as risk reducers, highlights the complex, multifactorial nature of food insecurity.

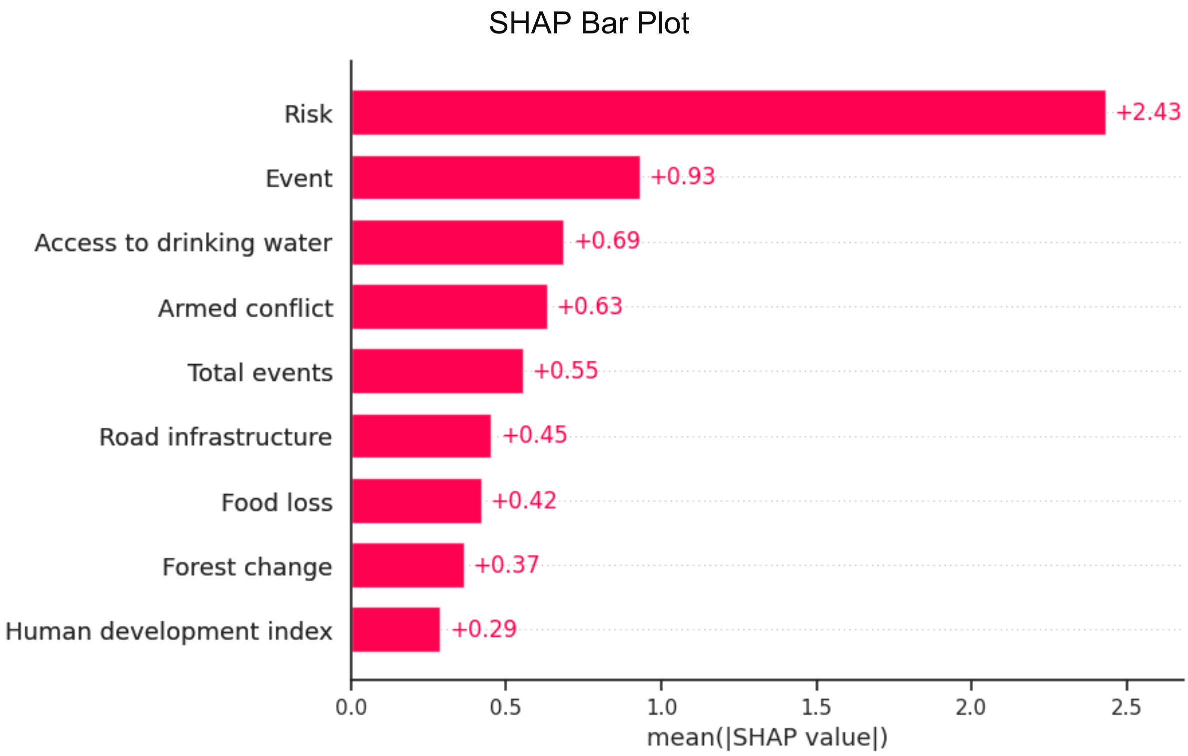

Figure 10 SHAP bar plot ranks the mean absolute SHAP values for each feature, illustrating their overall importance in the model's predictions. This plot highlights which features consistently play a significant role in the model's decision-making process.

The SHAP Bar Plot (Figure 10) gives an aggregate view of the overall importance of each feature across all countries. *Risk* stands out as the most influential feature, with a significantly larger mean SHAP value compared to other features. This suggests that addressing systemic risks, whether they be climate-related, economic, political, or all of them, would likely yield the largest reduction in global food insecurity. A similar conclusion was drawn in 2018 where they talked about the direct impact of one particular risk (climatic) and the impact it has in reducing the food insecurity in affected areas [21]. The fact that *Risk* dominates so strongly in this analysis indicates that it is a crucial variable for policy-makers to monitor and manage.

Event, *Access to drinking water*, and *Armed conflict* also have substantial mean SHAP values, indicating that they are critical drivers of food insecurity in many countries. The relatively high SHAP value for *Armed conflict* reflects the destabilising effect that conflict has on food systems. Countries embroiled in conflict are more likely to experience disruptions in food production, distribution, and access, often leading to acute food insecurity [6]. This reinforces the need for conflict resolution and peacebuilding efforts as part of broader food security strategies. Similarly, *Access to drinking water* and *Food loss* emerge as significant variables, highlighting the importance of infrastructure investments and reducing wastage in the food supply chain to enhance food security outcomes.

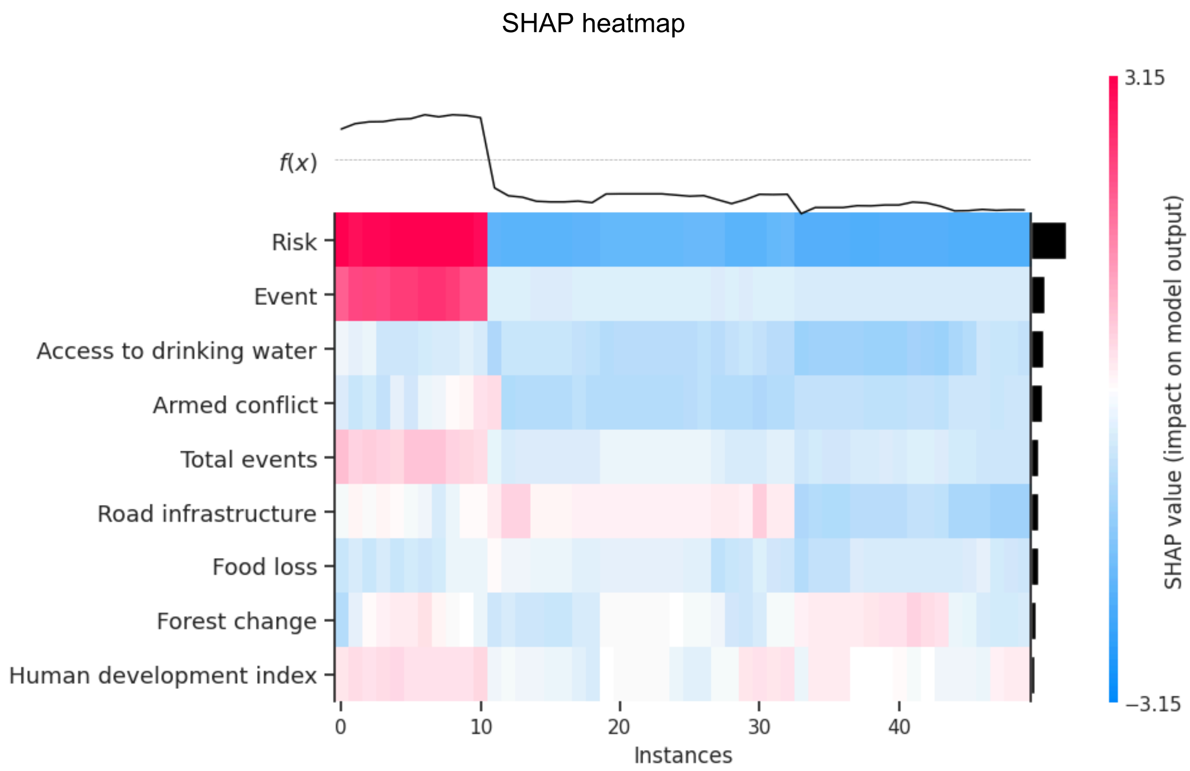

Figure 11. It visualises the impact of different features across multiple instances (countries) for predicting food assistance needs. The colour gradient represents SHAP values, with red indicating a strong positive impact on the prediction and blue indicating a strong negative impact.

In Figure 11 each column represents a country, and each row represents the features, with the colour of the cells indicating the direction and magnitude of the feature's contribution to food insecurity in that country. *Risk* is consistently red for many countries, highlighting that high risk levels are a major contributor to food insecurity across a wide range of contexts. The variation in the intensity of the red cells suggests that while risk is a key metric, its specific impact may differ from country to country depending on other interacting variables.

Event also shows a significant red cluster in many countries, particularly those that have experienced recent natural disasters, economic crises, or political upheaval. The variation in *Access to drinking water* across countries (both red and blue cells) again highlights its context-dependent role. In some countries, lack of water access is a key driver of food insecurity, while in others, better water access helps alleviate risk. This variability in SHAP values underscores the importance of tailored interventions based on the specific challenges that each country faces. For example, in some regions, improving infrastructure and access to clean water may yield the greatest benefits, while in others, addressing conflict or improving governance may be more effective.

Lastly, *Armed conflict* has a more uneven distribution in the heatmap, with certain countries showing deep red cells, indicating that conflict is a major contributor to their food insecurity. In contrast, countries with less internal or external conflict show neutral or even blue cells for this feature, suggesting that conflict resolution efforts could significantly reduce food insecurity in high-risk areas.

5. Challenges and Limitations

5.1. Data limitations

Although absolute certainty regarding the accuracy of the data is impossible, one can reasonably trust that the data collected by institutions such as the FAO, World Bank, and EIU provides a close approximation of reality. However, large-scale datasets used in this analysis are susceptible to systematic biases, reporting delays, or inconsistencies in measurement techniques. These limitations may introduce errors that can skew the findings, potentially leading to incorrect conclusions about the true extent of food insecurity. For instance, differences in methodologies between data sources or irregular data collection intervals could hinder the identification of critical patterns, particularly when certain variables, such as *Risk* or *Food loss*, depend heavily on reliable, consistent data collection over time.

Additionally, the nature of the cross-sectional data introduces temporal limitations. The data represents the state of variables on a yearly basis, namely from January 1 to December 31, yet the dynamics of food security can change rapidly within shorter time frames. Data collected toward the end of one year (e.g., December 2018) may be more closely related to the conditions of the following year (e.g., January 2019) than to the conditions from earlier in the same year. This lack of temporal granularity limits the ability to capture short-term fluctuations in food security and may obscure hidden trends that could emerge if time-series data were available. As such, the findings of this study are constrained by the temporal resolution of the data.

5.2. Model constraints

The Global Information and Early Warning System (GIEWS) alerts, which are central to this study's prediction of food insecurity, are issued at irregular intervals, making it difficult to compare countries over time. These alerts are often triggered by specific events or crises, which may not occur simultaneously across different regions. Consequently, the model is forced to rely on annual data for predictions, rather than generating monthly or weekly forecasts, which could provide more timely insights. This limitation restricts the model's ability to offer more granular predictions that could be crucial in responding to rapidly changing conditions.

Additionally, certain models like k-Nearest Neighbors and Logistic Regression struggled with the high-dimensional and non-linear nature of the food security data. While models such as Gradient Boosting and Random Forest performed well overall, the complexity of food security as a multifaceted issue—with contributing factors ranging from *armed conflict* to environmental degradation—means that even the best-performing models are constrained by the limits of available data and the model's capacity to capture these interactions. As seen in the Multi-layer Perceptron (MLP) model's performance, despite its ability to model non-linear relationships, it still underperformed in recall compared to ensemble methods, indicating that there are still limitations in fully capturing the intricate dependencies between the variables.

5.3. Future directions

To improve upon the current model and data constraints, future research could focus on leveraging Natural Language Processing (NLP) to analyse the specific causes of food insecurity as detailed in the GIEWS dataset. By extracting insights from the descriptive data accompanying each food insecurity alert, NLP techniques could help identify common patterns or triggers that lead to food crises, potentially revealing previously unrecognised factors that drive food insecurity following the approaches made by Molenaar et al. [29]. This could also enhance the model's ability to predict food insecurity by incorporating qualitative data into a more comprehensive analysis.

Another potential avenue for improvement is the use of time-series data, particularly when focusing on a single country or region. Transitioning from cross-sectional to time-series analysis would allow researchers to better capture the temporal dynamics of food insecurity, offering more accurate predictions by accounting for trends and recurring patterns. This could also enable more frequent, fine-grained predictions (e.g., monthly or weekly), which would provide policymakers with more actionable insights. Furthermore, analysing specific countries or regions in greater depth would allow for a more targeted understanding of localised factors that contribute to food insecurity, enabling more precise interventions.

6. Conclusion

The goal of this dissertation was to develop robust models to predict which countries require external assistance for food, utilising various socio-economic and environmental indicators such as *Risk*, *Access to drinking water*, *Food loss*, and *Armed conflict*. Through the application of machine learning techniques, this study sought to provide useful insights for policymakers to proactively address food insecurity and allocate resources efficiently. The results of this research demonstrate that while each machine learning model offers unique strengths and weaknesses, certain models clearly outperformed others in identifying countries at risk of food insecurity.

The Gradient Boosting and Random Forest models consistently showed the best performance, with both achieving high recall values (close to 0.98). These models were good at minimising false negatives, which is crucial in the context of food security, as failing to identify a country in need of assistance could have devastating humanitarian consequences. The Gradient Boosting model, in particular, demonstrated a strong balance between minimising false negatives and maintaining a high true positive rate, as reflected in the confusion matrix where only 3 false negatives were recorded. This precision is essential for ensuring that resources are directed toward countries that genuinely require aid, without wasting them on misclassified cases.

However, it is important to acknowledge that there is no single “best” model, as the choice of model depends heavily on the specific goals and constraints of the task at hand. While Gradient Boosting and Random Forest offered the highest performance in terms of recall and accuracy, their complexity and lower interpretability may not always be ideal in every scenario. In contrast, Decision Trees, though slightly less accurate, provide the distinct advantage of interpretability. Decision Trees allow for a transparent understanding of the model’s decision-making process, which is invaluable in policy-driven environments where transparency and accountability are critical. Policymakers need to understand why a particular prediction was made, and Decision Trees can provide that clarity.

Other models, such as k-Nearest Neighbors (kNN) and Multi-layer Perceptron (MLP), performed reasonably well but showed limitations in certain areas. The kNN model exhibited a higher rate of false positives, which could lead to the inefficient allocation of resources. The MLP, despite its ability to model non-linear relationships, showed a slight decrease in recall compared to the ensemble methods, suggesting that it might miss a subset of countries in need. Moreover, both Logistic Regression and Support Vector Machines (SVM) underperformed in comparison to the ensemble methods due to their limited ability to capture the complex, non-linear interactions present in the data. These models struggled to handle the intricate relationships between variables such as *Land degradation*, *Forest change*, and *Human Development Index*, and as a result, their recall values were significantly lower, making them less suitable for this particular study.

The SHAP (SHapley Additive exPlanations) analysis further showed the importance of individual features in predicting food insecurity. *Risk*, *Access to drinking water*, and *Armed conflict* emerged as the most influential variables driving food insecurity. SHAP analysis helped quantify the contribution of these features, providing actionable insights into how each factor affects food security. For example, countries with low access to potable water and higher exposure to risks such as natural disasters or conflict are more likely to require external assistance, which aligns with real-world expectations. The use of SHAP values enhances the interpretability of the models, offering politicians, businesspeople and investors a clearer understanding of which factors are most critical in driving the need for food assistance.

Despite the strong performance of the models, the study may have certain data limitations. The reliance on cross-sectional data (capturing information only at specific points in time) limited the ability to track dynamic changes in food security over shorter time frames. For instance, data collected at the end of one year may be more closely related to the conditions of the following year, which introduces challenges when trying to predict outcomes based on annual data. Furthermore, the irregular timing of the GIEWS alerts made it difficult to perform fine-grained predictions, such as monthly or weekly forecasts, restricting the model's ability to provide timely responses to emerging crises. Future research should explore the use of time-series data to improve the temporal resolution of predictions and provide more granular insights into food security trends.

Additionally, future research could benefit from leveraging Natural Language Processing (NLP) techniques to analyse the specific causes of food insecurity as described in the GIEWS alerts. By extracting qualitative information from textual descriptions, NLP could uncover patterns that are not captured in structured data, enhancing the predictive power of the models. Moreover, focusing on time-series data for specific regions or countries would allow for more accurate, real-time predictions that could better inform immediate interventions.

7. References

[1] Alkama, Ramdane, and Alessandro Cescatti. "Biophysical climate impacts of recent changes in global forest cover." *Science* 351.6273, 2016: 600-604.

[2] Amin, Modhurima Dey, Syed Badruddoza, and Jill J. McCluskey. "Predicting access to healthful food retailers with machine learning." *Food Policy 99,* 2021: 101985.

[3] Assembly, General. "Resolution adopted by the General Assembly, 6 July 2017." *Technical Report* A/RES/71/313.

[4] Backer, David, and Trey Billing. "Validating famine early warning systems network projections of food security in Africa, 2009–2020." *Global Food Security 29*, 2021: 100510.

[5] Broussard, Nzinga H., and Sharad Tandon. "Food Insecurity Measures: Experience-Based Versus Nutrition-Based Evidence From India, Bangladesh, and Ethiopia." *AgEcon*, 2016, .10.22004

[6] Brück, Tilman, et al. "The relationship between food security and violent conflict." *Report to the Food and Agriculture Organization*, ISDC, Berlin, 2016.

[7] Buchanan-Smith, Margaret, and Susanna Davies. *Famine early warning and response: the missing link.* London, 1995.

[8] Caffo, Brian. *"Regression models for data science in R." A companion book for the Coursera Regression Models class,* 2015.

[9] Calder, Ian R. *"*Forests and water—ensuring forest benefits outweigh water costs.*" Forest ecology and management 251.1-2,* 2007: 110-120.

[10] Chamara, R. M. S. R., et al. *"Role of artificial intelligence in achieving global food security: a promising technology for future."* Sri Lanka Journal of Food and Agriculture 6.2, 2020.

[11] Chawla, Nitesh V., et al. "SMOTE: synthetic minority over-sampling technique." *Journal of artificial intelligence research 16,* 2002: 321-357.

[12] Deléglise, Hugo, et al. "Food security prediction from heterogeneous data combining machine and deep learning methods.*" Expert Systems with Applications 190*, 2022: 116189.

[13] Devereux, Stephen. "Food insecurity and famine.*" Handbook of African Development. Routledge,* 2018. 183-201.

[14] Dong, Guozhu, and Huan Liu, eds. Feature engineering for machine learning and data analytics. *CRC press,* 2018.

[15] Ellison, David, et al. "Trees, forests and water: Cool insights for a hot world." *Global environmental change 43,* 2017: 51-61.

[16] FAO, GIEWS. "GIEWS-Global Information and Early Warning System." 2020.

[17] Gomiero, Tiziano. "Soil degradation, land scarcity and food security: Reviewing a complex challenge." *Sustainability 8.3,* 2016: 281.

[18] Grant, Richard, et al. "Climate adaptation and resilience indices for the Caribbean region: an assessment of four leading indices." *Climate and Development,* 2023: 1-11.

[19] Han, Hui, Wen-Yuan Wang, and Bing-Huan Mao. "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning." *International conference on intelligent computing.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[20] Harris, Jabez Jay. *"A machine learning approach to forecasting consumer food prices."* 2017.

[21] Hasegawa, Tomoko, et al. "Risk of increased food insecurity under stringent global climate change mitigation policy." *Nature climate change 8.8,* 2018: 699-703.

[22] Huhtamaa, Heli. Frosts, floods and famines: climate in relation to hunger in north-east Europe AD 1100-1550. *MS thesis. Itä-Suomen yliopisto,* 2011.

[23] Irshad, Muhammad Tausif, et al. "SenseHunger: Machine learning approach to hunger detection using wearable sensors." *Sensors 22.20,* 2022: 7711.

[24] Ishangulyyev, Rovshen, Sanghyo Kim, and Sang Hyeon Lee. *"Understanding food loss and waste—why are we losing and wasting food?." Foods 8.8,* 2019: 297.

[25] James, Gareth, et al. *An introduction to statistical learning: With applications in python.* Springer Nature, 2023.

[26] Lotfi, Reza, et al. "A robust and resilience machine learning for forecasting agri-food production." *Scientific Reports 12.*1 , 2022: 21787.

[27] Masset, Edoardo. "A review of hunger indices and methods to monitor country commitment to fighting hunger." *Food policy 36,* 2011: S102-S108.

[28] Messer, Ellen, Marc J. Cohen, and Thomas Marchione. *"Conflict: A cause and effect of hunger.",* 2001: 12-15.

[29] Molenaar, Annika, et al. "Using Natural Language Processing to Explore Social Media Opinions on Food Security: Sentiment Analysis and Topic Modeling Study." *Journal of Medical Internet Research 26,* 2024: e47826.

[30] Molnar, Christoph. *Interpretable machine learning.* Leanpub book, 2020.

[31] Murray, Christopher JL, et al. "Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019." *The lancet 396.10258,* 2020: 1223-1249.

[32] Okori, Washington, and Joseph Obua. "Machine learning classification technique for famine prediction." *Proceedings of the world congress on engineering.* Vol. 2. No. 1. 2011.

[33] Okori, Washington, and Joseph Obua. "Supervised learning algorithms for famine prediction." *Applied Artificial Intelligence 25.9,* 2011: 822-835.

[34] Owen, Louis. Hyperparameter Tuning with Python: Boost your machine learning model's performance via hyperparameter tuning. *Packt Publishing Ltd,* 2022.

[35] Oxford Analytica. "World hunger is likely to rise as 2030 UN target nears." *Emerald Expert Briefings oxan-db,* 2019.

[36] Ritchie, Hannah, Fiona Spooner, and Max Roser. "Clean water." *Our world in data,* 2023.

[37] Soleh, Agus Mohamad, and Budi Susetyo. "Rotation Double Random Forest Algorithm to Predict The Food Insecurity Status of Households." *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) 8.1,* 2024: 33-41.

[38] Swift, Jeremy. "Understanding and preventing famine and famine mortality." *IDS bulletin 24* 1993: 1-1.

[39] United Nations. “Human Development Index.”*Nations Development Programme,* 2024, https://hdr.undp.org/data-center/human-development-index#/indicies/HDI.

[40] Utuk, Ime O., and Ekong E. Daniel. "Land degradation: a threat to food security: a global assessment." *J Environ Earth Sci 5.8*, 2015: 13-21.

[41] Verdin, James, et al. "Climate science and famine early warning." *Philosophical Transactions of the Royal Society B: Biological Sciences 360.1463,* 2005: 2155-2168.

[42] Viechtbauer, Wolfgang, and Mike W‐L. Cheung. "Outlier and influence diagnostics for meta‐analysis." *Research synthesis methods 1.2,* 2010: 112-125.

[43] Vos, Rob, Brendan Rice, and Nicholas Minot. *"2022 Global report on food crises: Joint analysis for better decisions.",* 2022.

[44] Wade, Corey, and Kevin Glynn. Hands-On Gradient Boosting with XGBoost and scikit-learn: Perform accessible machine learning and extreme gradient boosting with Python. *Packt Publishing Ltd,* 2020.

[45] Westerveld, J. J. L. Modelling Food Insecurity in Ethiopia: Towards a machine learning model that predicts the transitions in food security using scalable features. *MS thesis, Utrecht University.* 2019.

[46] Wikipedia, "Linear interpolation." *Wikimedia Foundation,* 18 Sep. 2024. *https://en.wikipedia.org/wiki/Linear_interpolation#:~:text=In%20mathematics%2C%20linear%20interpolation%20is,set%20of%20known%20data%20points.*

[47] World Bank Group. *"World Bank Group Data"* , Sep 2024, https://www.worldbank.org/en/home