Optimizing Food Baskets

For Nigerian Households

Presented by Pearry Khachichit, Jessie Lee, Rahila Sule, and Mimi Vinaiphat

AGENDA

1. PROBLEM STATEMENT
2. DATA EXPLORATION
3. METHODS
4. RESULTS & RECOMMENDATION

PROBLEM STATEMENT

"With a population of over *200 million people* in Nigeria, the WFP's hunger map reports **55.4 million people** with **insufficient food consumption** – that's over a quarter of the population living with food uncertainty."

-World Food Programme-

Photo by Sheyi Owolabi on Unsplash

Research Question

How to design a ***nutritionally adequate***, ***low-cost*** and ***culturally acceptable*** food basket* for Nigerian households based on similarity in coping strategies?

*Food basket: list of locally available foods

OBJECTIVES

1) Calculate Nigerian households' food insecurity levels based on the Coping Strategies Index (CSI)
2) Create sub-groups of Nigerian households with similar coping patterns
3) Predict Nigerian household food insecurity levels
4) Design **optimal food baskets** for Nigerian households in fulfillment of U.S. FDA nutrient recommendations considering *affordability* and *cultural acceptability*

DATA EXPLORATION

Data Summary

Nigeria General Household Survey, Panel 2018-2019 (Wave 4)

Implemented in collaboration with the World Bank

- **Data type:** Survey of approximately **5,000 households** representative of the six geopolitical zones and all 37 states in Nigeria
- **Unit of analysis:** Households, Communities
- **Scope:** The survey covered various topics under 3 main questionnaires:
  1) Household*: demographic | economic shocks | food price & expenditure
  2) Agriculture: crop farming and livestock rearing
  3) Community*: socio-economic information

COPING STRATEGIES INDEX (CSI)

An indicator of household food insecurity

The CSI assesses the behaviour of people when they cannot access enough food. CSI is calculated based on a series of questions aimed at providing different answers to the sole question: *“What do you do when you don’t have adequate food and don’t have the money to buy any?”*

Uses of CSI :

- Quick status indicator of the extent of food insecurity for decision-making & interventions
- Early warning indicator of food insecurity

Method:

The CSI is calculated as a weighted sum of quantified coping behaviors (i.e. yes/no responses to assessment questions)

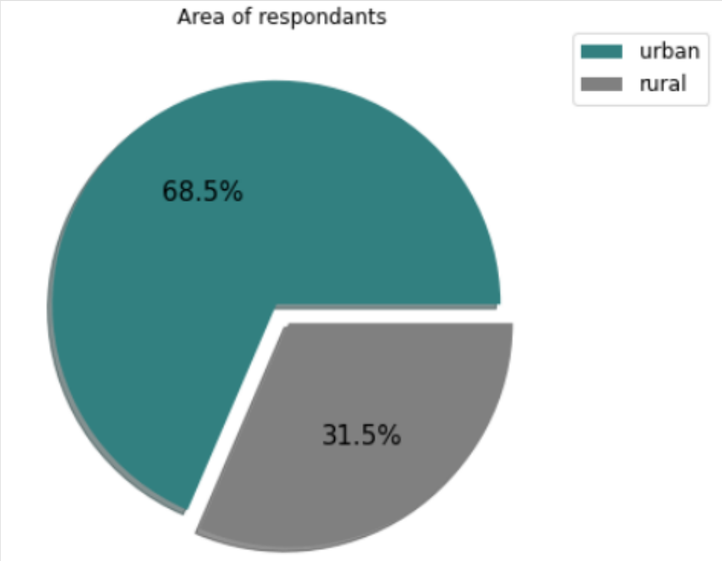
CSI Feature Engineering

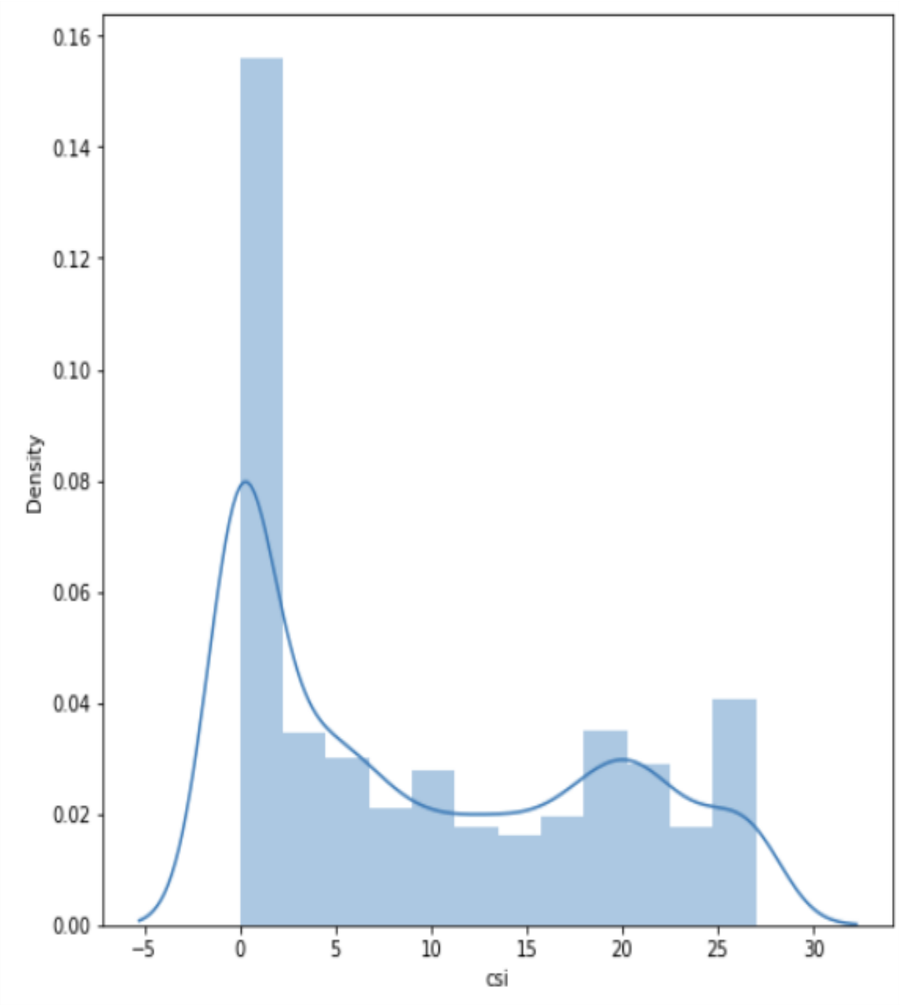
No.	Questions	Severity Weight
1	Household worried about not having enough food to eat in the last 30 days	2
2	Household unable to eat healthy and nutritious/preferred food in the past 30 days	3
3	Household ate only a few kinds of food in the last 30 days	1
4	Household had to skip a meal in 30 days	4
5	Household ate less that you thought you should in the last 30 days	1
6	Household ran out of food in the last 30 days	3
7	Household were hungry but did not eat in the last 30 days	3
8	Household went without eating for a whole day in the last 30 days	4
9	Household restricted consumption in order got children to eat in the last 30 days	2
10	Household borrowed food, or relied on help from a friend/relative in the last 30 days	2
11	Household been faced with a situation where they did not have enough food	2

Max score: 27

Weighted score ranges: 0-9(group 1), 10-18 (group 2), 19-27(group 3) (higher is worse)

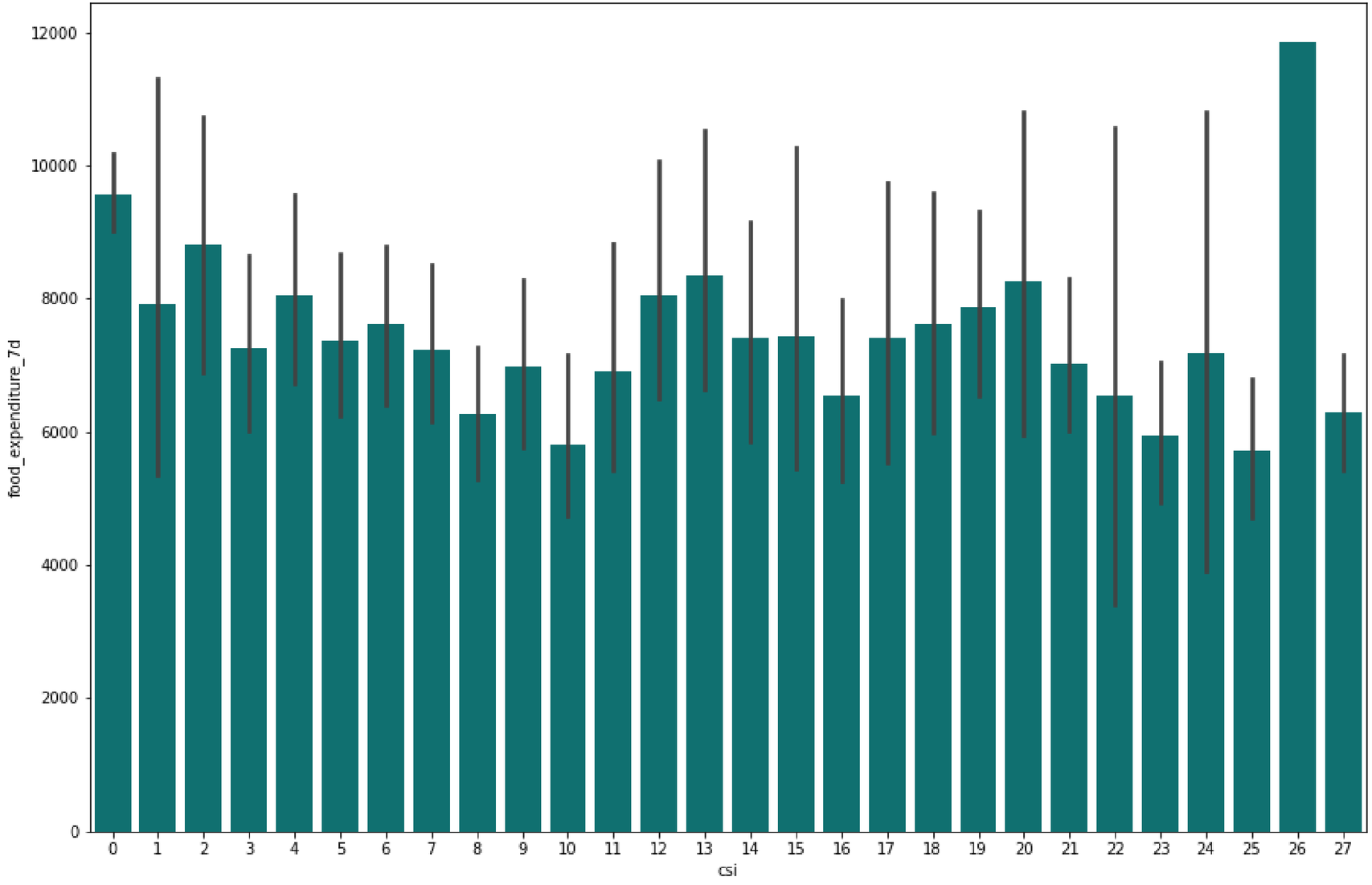
EXPLORATORY DATA ANALYSIS





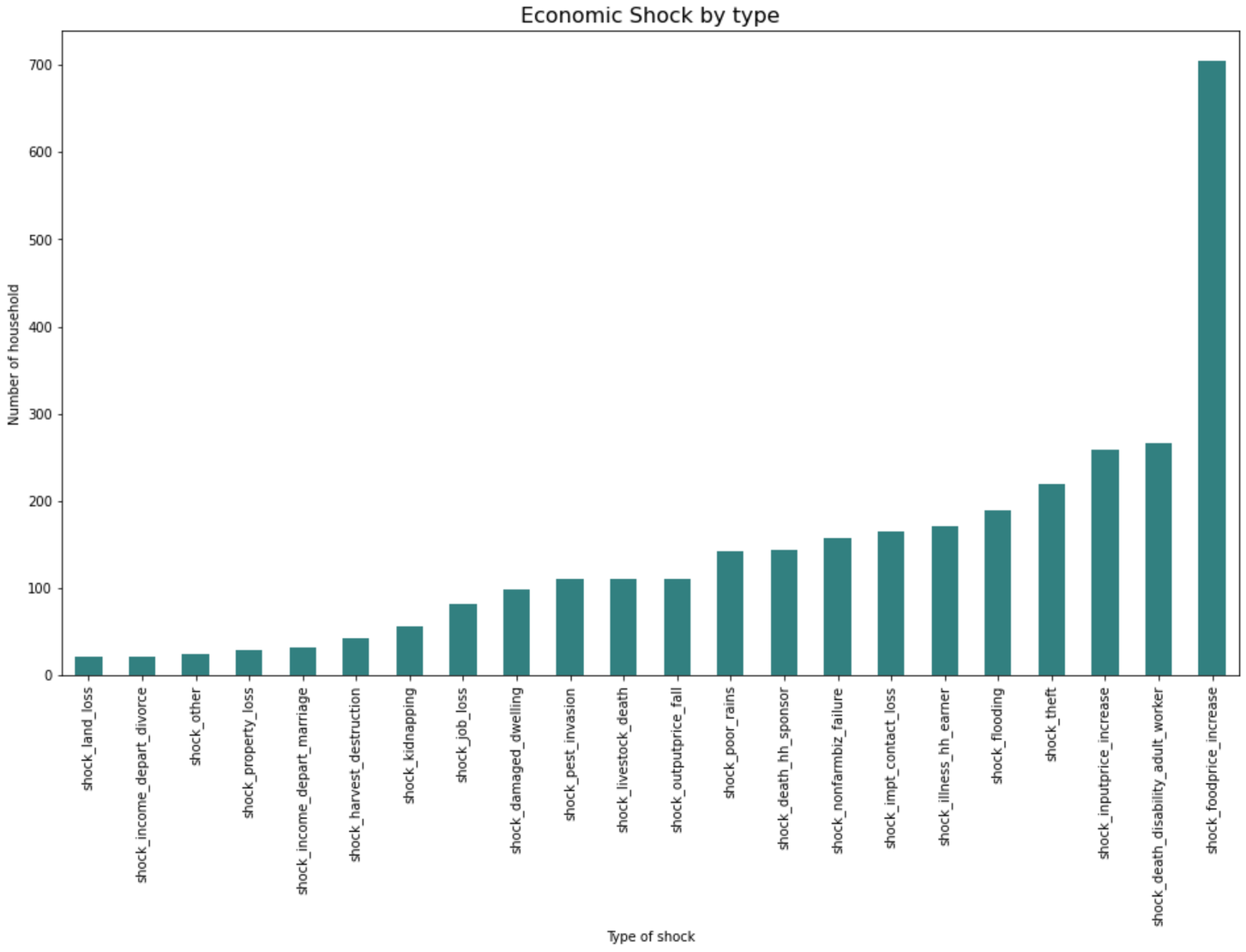
- About 68% of respondent households live in urban areas
- Based on the CSI about half of respondent households are potentially facing high food insecurity, on average

EXPLORATORY DATA ANALYSIS



- From this, the total food spending does not seem to reflect with the CSI score. You would imagine that those with higher score would have lower spending as they might not have enough food
- The error bars indicate the deviations and each CSI's bar are quite fluctuate as well.

EXPLORATORY DATA ANALYSIS



- Based from the visualization, household got affected most by the increase in food price, followed by the member's disability. This can be implied that major concern is regarding with food price.

METHODS

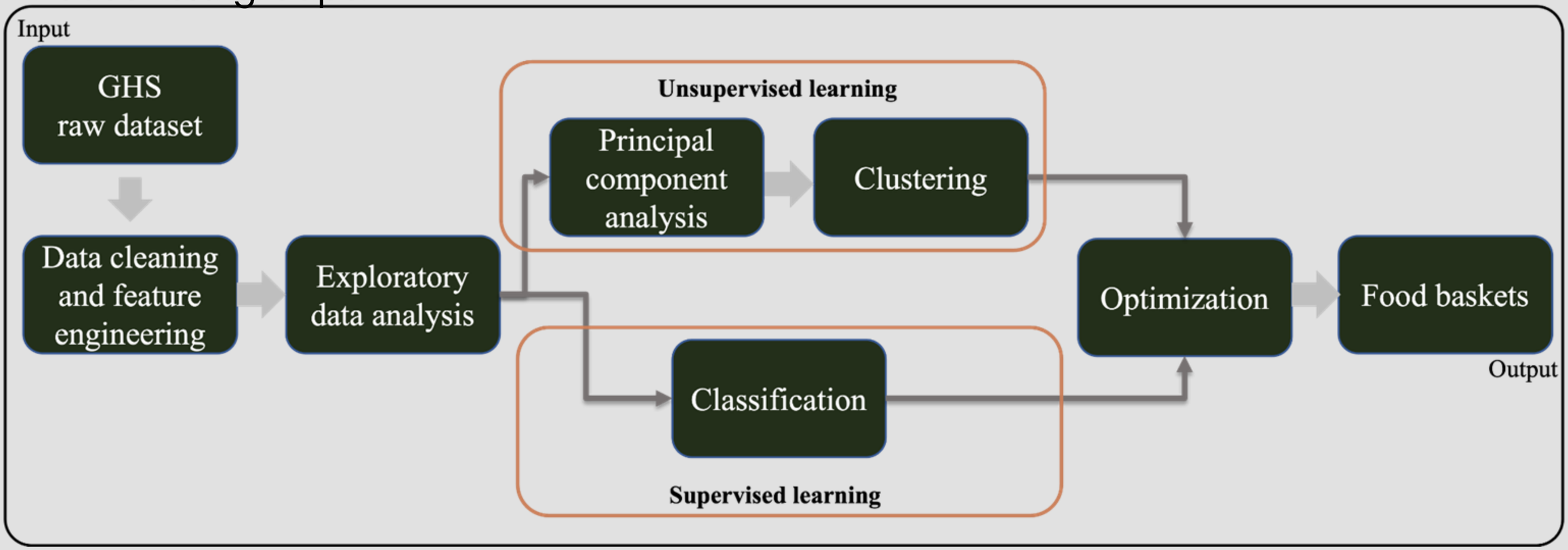
Assumptions

The project is subject to the following assumptions:

- Reference family size for computations is four adult household members for each household
- Total food expenditure in dataset for each household is their average weekly food expenditure
- Maximum food basket cost is based on median household expenditure for each group/cluster
- Consumption coping behavior questions used for computing CSI are relevant to the Nigerian cultural context
- Household members do not require specific dietary requirements
- Food prices are converted to per kilogram for solid items and per litre for liquid items
- Food prices are captured based on the average food prices for each item
- Food nutrients for each food item are converted based on the same unit as the food price (per kg for solid items and per litre for liquid items)

Model Plan

The methodology was first to develop food insecurity profiles for Nigerian households using machine learning, then subsequently, create optimized food baskets for the household groups.



MACHINE LEARNING

Machine Learning Objective

The machine learning goal was to develop food insecurity profiles for Nigerian households based on the computed Coping Strategies Index(CSI) and the following features:

- household sector (i.e. urban vs rural)
- household dwelling type
- loan access
- total weekly food expenditure
- percentage of food consumed by sources (self-produced | purchased | gifts)
- household economic shocks

Supervised Learning

Use case

Use the supervised learning models to predict the food insecurity level of households whose CSI information is unavailable.

Goal

The supervised learning model was aimed at predicting the food insecurity levels of Nigerian households. Thus, the true labels of the dependent variable were based on the households' food insecurity groups: Low = 1, Medium = 2, High = 3

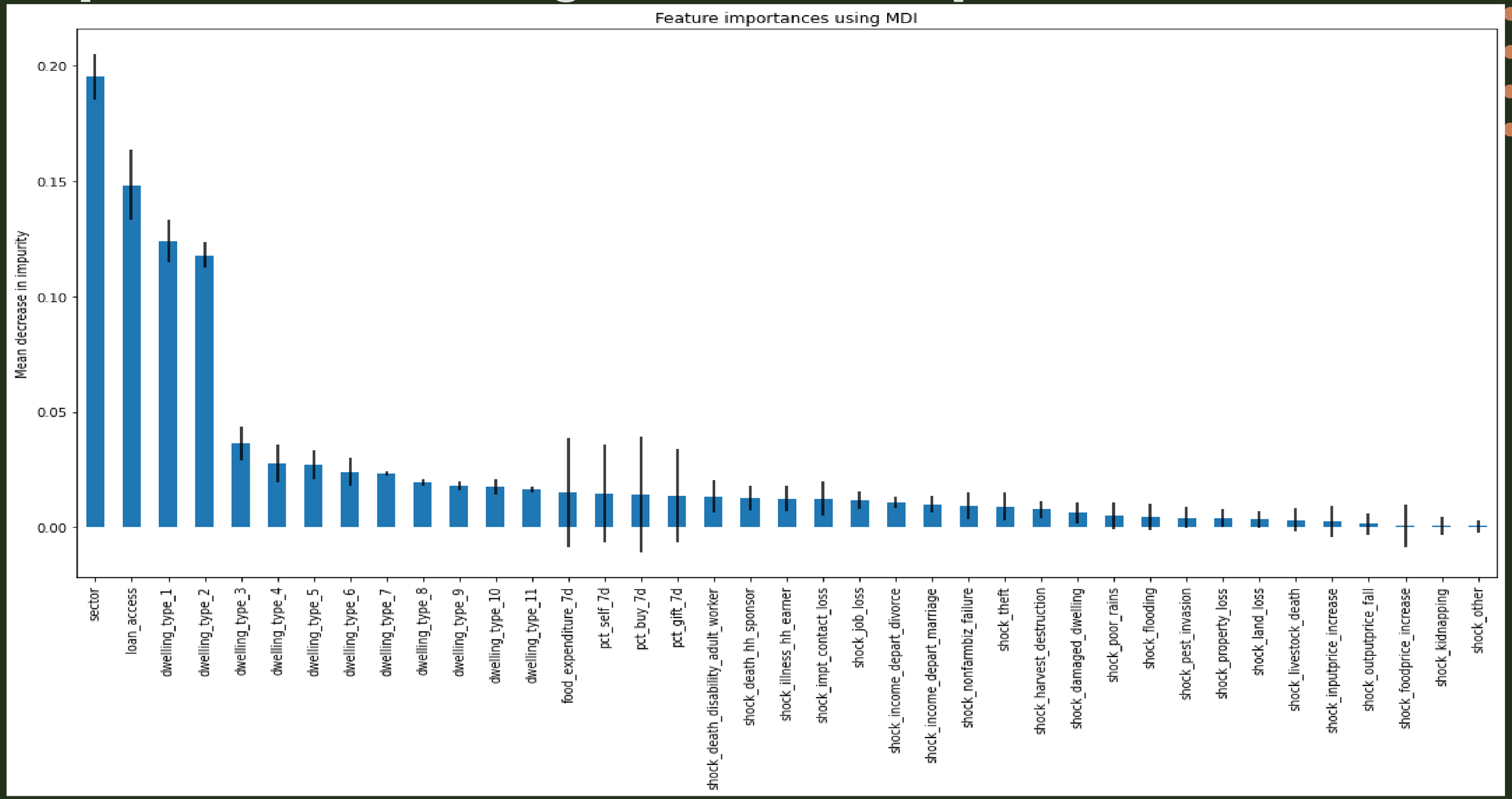
Models

The following models were evaluated for highest performance based on accuracy:

Model	Train Accuracy
Linear Discriminant Analysis	57.3%
Quadratic Discriminant Analysis	56.9%
K-Nearest Neighbors	56.3%
Decision Tree	57.1%
Random Forest	**57.4%**

Test Accuracy: ~59%

Supervised Learning – Feature Importance



Unsupervised Learning

Use Case

Cluster similar households when all the households' CSI information is available.

Goal

The unsupervised learning model was aimed at identifying homogeneous subgroups within Nigerian households at similar food insecurity severity levels.

Feature

feature used in the supervised learning model + CSI

Models

- K-Means **Best K = 2**
- Gaussian Mixture Model

$$s = \frac{b - a}{max(a, b)}$$

a: Mean distance between a point and all other points in the same cluster
b: Mean distance between a point and all other points in its nearest cluster

Average Silhouette Coefficient Score

Median weekly food expenditure

Random Forest

Supervised Model

- Low: NGN 5,170
- Medium: NGN 3,660
- High: NGN 2,890

K-means (k=2)

Unsupervised Model

- Cluster 1: NGN 4,270
- Cluster 2: NGN 6,860

OPTIMIZATION

Low-cost food basket

Input parameters:

- Food prices
- Nutrient composition for each food item
- Median food expenditure for each household group

Decision variables: Optimized weight of food item in kilograms in the food basket

Objective: Minimize weekly total food expenditure

Constraints:

- Total nutrients of food items in the food basket should at least meet the daily reference values recommended by the U.S. FDA
- Total cost of food items should not exceed median household food expenditure for the household group

Low-cost food basket

Problem formulation

Indices:

n: Total number of food items
m: Total number of food nutrient groups

$i: 1, \dots, n$ food items
$j: 1, \dots, m$ food nutrition component

Parameters:

E: median weekly food expenditure
c_i: cost of food item i
D_j: Daily reference value for food nutrient j
N_{ij}: weight of food nutrient j in food item i

Decision variables:

f_i: weight of food item i to have in food basket

Objective function:

$$Minimize \sum_{i=1}^{n} c_i * f_i$$

Subject to:

Food expenditure constraint

$$\sum_{i=1}^{n} c_i * f_i \leq E$$

Non-negativity constraint

$$f_i \geq 0 \; \forall \; i = 1, \dots, n$$

Nutrient constraint

$$\sum_{i=1}^{n} N_{ij} * f_i \geq D_j * 4 * 7 \; \forall \; j = 1, \dots, m$$

Low-cost food basket optimized for cultural acceptability

Input parameters:

- Food prices
- Nutrient composition for each food item
- Median food expenditure for each household group

Decision variables: Optimized weight of food item in kilograms in the food basket

Objective: Minimize total squared relative deviation of the optimized food basket from the reported food intake

Constraints:

- Total nutrients of food items in the food basket should at least meet the daily reference values recommended by the U.S. FDA
- Total cost of food items should not exceed median household food expenditure for the household group

Low-cost food basket optimized for cultural acceptability

Problem formulation

Indices:

n: Total number of food items
m: Total number of food nutrient groups

$i: 1, \dots, n$ food items
$j: 1, \dots, m$ food nutrition component

Parameters:

F_i: average weight of same food item i reported to be consumed in a week
E: median weekly food expenditure
c_i: cost of food item i
D_j: Daily reference value for food nutrient j
N_{ij}: weight of food nutrient j in food item i

Decision variables:

f_i: weight of food item i to have in food basket

Objective function:

$$Minimize \sum_{i=1}^{n} \left(\frac{f_i - F_i}{F_i}\right)^2$$

Subject to:

Food expenditure constraint

$$\sum_{i=1}^{n} c_i * f_i \leq E$$

Non-negativity constraint

$$f_i \geq 0 \; \forall \; i = 1, \dots, n$$

Nutrient constraint

$$\sum_{i=1}^{n} N_{ij} * f_i \geq D_j * 4 * 7 \; \forall \; j = 1, \dots, m$$

RESULTS & RECOMMENDATION

Results

Machine Learning Model	Optimization Model	Household group	Cost (NGN)	Number of food items
Supervised learning	Low-cost food basket	Low food insecurity	115	4
		Medium food insecurity	115	4
		High food insecurity	115	4
	Low-cost food basket optimized for cultural acceptability	Low food insecurity	5,170	72
		Medium food insecurity	3,660	54
		High food insecurity	2,890	51
Unsupervised learning	Low-cost food basket	Cluster 1	134	5
		Cluster 2	134	5
	Low-cost food basket optimized for cultural acceptability	Cluster 1	4,270	67
		Cluster 2	6,860	78

Policy Recommendation

Stakeholders:

- Nigerian Ministry of Health
- National Agency for Food and Drug Administration and Control

Recommendations:

- Adopt optimized food basket model for hunger intervention programs aimed at reducing the food insecurity burden for Nigerian households
- Incorporate estimated food basket cost as a critical component when determining transfer values for cash transfer programs

Limitations

Data	Non-standard measurement units
	Food baskets should be based on data from national food intake surveys
	Binary responses instead of frequency for coping questions
	Only one week’s food expenditure available
Analytical	Limited food nutrients are considered
	Possible measurement errors due to survey response error
	Cost calculated is only the cost of food items
Economic	Seasonal fluctuations and volatility of food prices

Future Work

- Consider more features in machine learning related to food access
  e.g. nearness to markets or supermarkets, availability of food items in community
- Incorporate Recommended Nutrient Intake (RNI),
  Estimated Energy Requirements (EER) and utilize macronutrient distribution range
- Food basket for different household members and sectors
- Incorporate more food items and create food groups with weights to ensure energy giving and nutritious food items receive higher priority in food basket inclusion
- Switch from a deterministic to stochastic optimization model to simulate different food basket designs based on differing household weekly food expenditures

THANK YOU

Questions?

Photo by Zoe Schaeffer on Unsplash