**CHAPTER FOUR**

**IMPLEMENTATION**

**4.1 INTRODUCTION**

Hunger remains a critical global challenge, affecting millions of lives across the world. An extensive analysis using Python was conducted in order to gain insights into the factors contributing to hunger and food insecurity. This report presents a detailed overview of the analysis, data visualization and key findings.

The primary objective is to understand the dynamics of hunger prevalence, affordability of a healthy diet and food prices. By analyzing these factors, trends, correlations and potential areas for intervention in the fight against global hunger can be identified.

**4.2 Data Sources**

**Hunger Data:** the primary dataset was gotten from 'Hunger and Undernourishment Data-1.xlsx,' which contains information on the prevalence of undernourishment across countries and years.

**Affordability Data:** Data on the affordability of a healthy diet is sourced from 'Cost and Affordability of a Healthy Diet (CoAHD)-1.xlsx.'

**Employment Data:** 'employment indicator.csv' provides insights into employment trends, which can influence access to food.

**Food Price Data**: 'Food Prices cost.csv' contains consumer prices, including food indices and food price inflation.

Before conducting the analysis, a comprehensive data preprocessing was performed to ensure data quality and consistency.

**4.3 Data Cleaning and integration**

The following step was followed in order to clean and integrate the dataset:

* Irrelevant columns were removed from the datasets.
* Missing values were handled through appropriate methods such as imputation or removal.
* Outliers were identified and addressed to prevent skewing the analysis.

To enable a holistic analysis, datasets based on common columns like 'Country' and 'Year' was merged. This allowed the combine of hunger data, affordability data, employment data, and food price data into a unified dataset.

**4.4 Exploratory Data Analysis (EDA)**

The analysis begins with exploratory data analysis to gain initial insights into the datasets.

**4.4.1 Hunger Analysis for highest prevalence undernourishment**

The top 20 countries with the highest prevalence of undernourishment in 2001 was identified. This helps pinpoint regions where hunger was most severe and visualize the data using a bar chart.

Pandas was used to filter the dataset specifically for the year and then the top selection of countries with the highest prevalence of undernourishment. The Seaborn library was then employed to create a bar chart where the x-axis represents the prevalence of undernourishment and the y-axis lists the countries. Each bar in the chart signifies a country and its height corresponds to the prevalence of undernourishment in that country.

The resulting bar chart provides a clear visual representation of the countries with the highest prevalence of hunger in the year 2001 clearly sowing that Angola tops the list on the plot and Somalia being the highest in the year 2019. This analysis helps identify regions or nations where hunger was most acute during that period. Understanding the specific countries facing severe hunger issues is essential for targeting interventions and aid effectively.

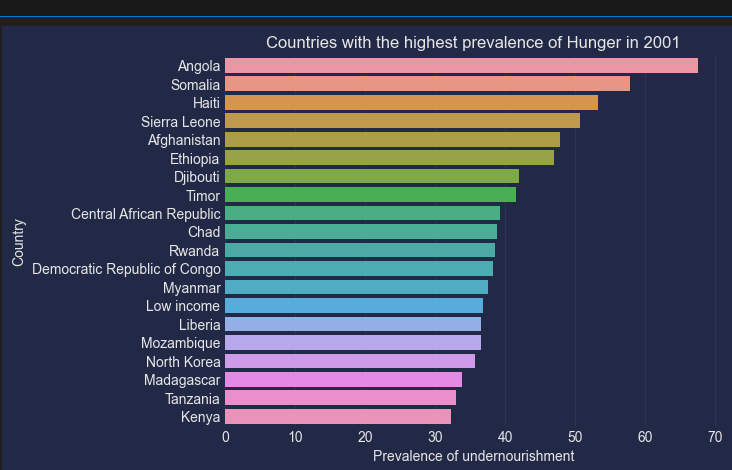


Fig 4.4.1: Bar chart showing countries with highest prevalence of hunger in 2001

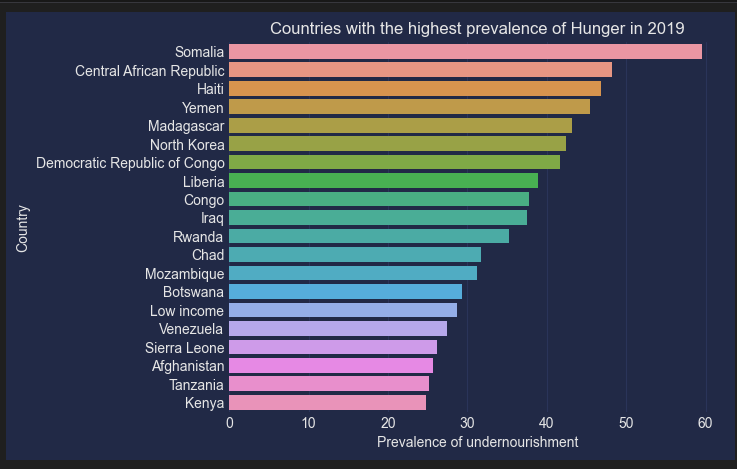


Fig 4.4.1: Bar chart showing countries with highest prevalence of hunger in 2019

**4.4. 2 Hunger Analysis for lowest hunger prevalence undernourishment**

Understanding and visualizing the countries with the lowest prevalence of undernourishment in the year 2001 was delved into.

The first step in our analysis is to identify and isolate the countries with the lowest hunger prevalence in the year 2001. To accomplish this, the Pandas library was used by filtering the dataset to include only the data for the year 2001 and then select the 20 countries with the smallest prevalence of undernourishment. This dataset now contains the countries that experienced the least hunger-related issues during that specific year. Once the selected data are gotten, seaborn a powerful data visualization library is employed to visualize the bar chat. The x-axis represents the prevalence of undernourishment while the y-axis displays the list of countries with each bar in the chart signifying a country and the height of the bar corresponds to the prevalence of undernourishment in that country.

The resulting bar chart provides a highly informative visual representation showing Australia, Austria, Belarus, Belgium and others with the lowest prevalence of hunger in the year 2001 and Algeria inclusively for year 2019. This analysis aids in identifying regions or nations where hunger was relatively less acute during that period. Understanding which countries fared better in terms of food security in a specific year is invaluable for policymakers and organizations working to combat hunger as it enables them to learn from successful strategies and allocate resources efficiently.

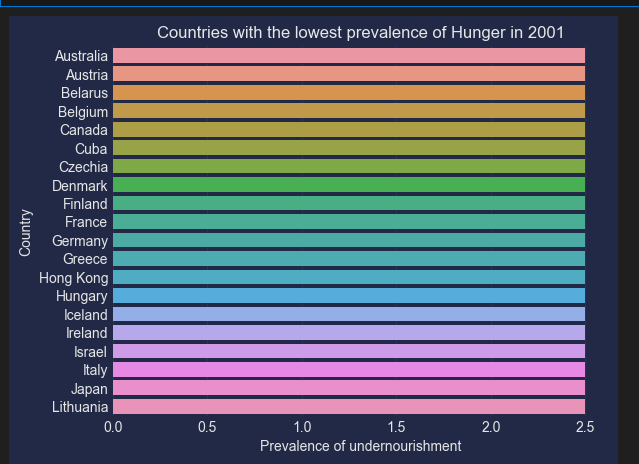


Fig 4.4.2: Bar chart showing countries with lowest prevalence of hunger in 2001

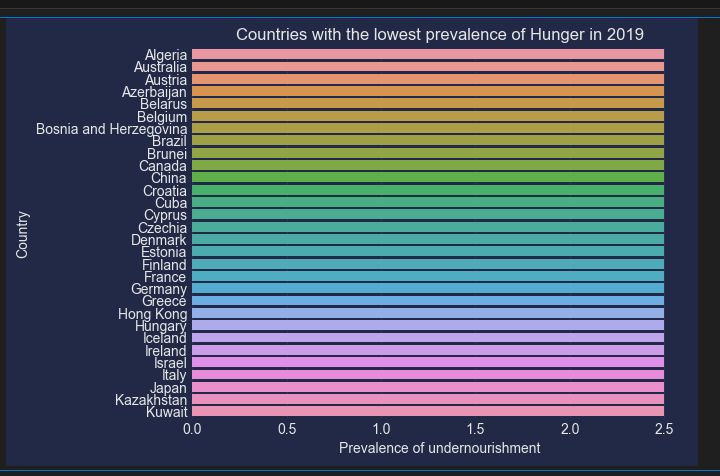


Fig 4.4.2: Bar chart showing countries with lowest prevalence of hunger in 2019

**4.4.3 Analyzing Changes in Hunger Prevalence from 2001 to 2019**

Understanding how the prevalence of undernourishment has changed in various countries from the year 2001 to 2019 was focused on.

**Calculating Prevalence Changes:** Quantifying the changes in hunger prevalence, the difference between the prevalence values in 2001 and 2019 was calculated. The dataset is grouped by country and for each country, the first available data point was subtracted (from 2001) from the last available data point (from 2019). The resulting diff\_df DataFrame now contains the computed differences providing insights into how each country's hunger situation has evolved over this 18-year period.

Once these differences have been calculated, we proceed to visualize the data using a bar chart created with Seaborn. The x-axis represents the prevalence of undernourishment changes (the difference between 2019 and 2001) while the y-axis displays the list of countries.

The resulting bar chart provides a clear and insightful view of the countries that have made the most significant strides in reducing hunger prevalence from 2001 to 2019 and vice versa with Yemen having the highest rise prevalence of hunger and Angola being the top of the biggest reduction in prevalence of hunger. Countries with the highest bars have shown the most substantial improvement indicating successful efforts in addressing food security and reducing hunger and the other way round.

This analysis is of utmost importance for policymakers, as it highlights success stories and strategies that have effectively reduced hunger in specific regions. Identifying and learning from these countries can aid others in implementing similar strategies to combat hunger.

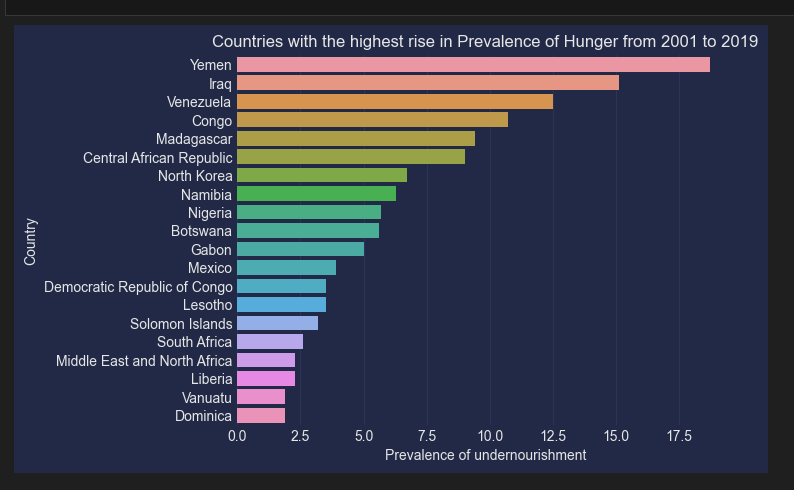


Fig 4.4.3: Bar chart showing countries highest rise prevalence of hunger between 2001- 2019

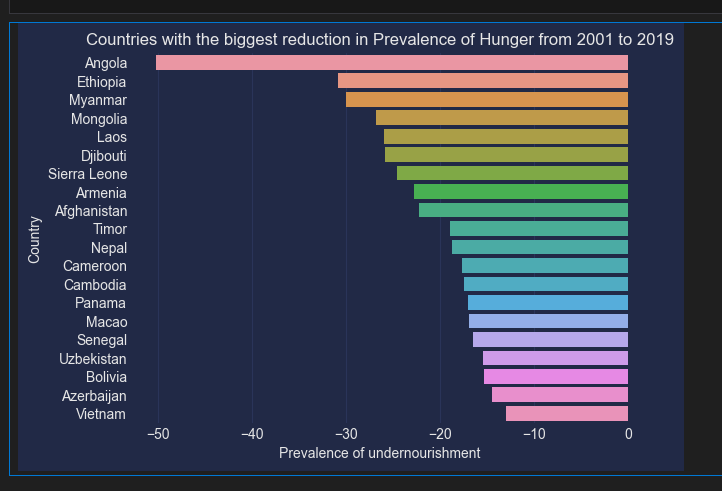


Fig 4.4.3: Bar chart showing countries with biggest reduction in prevalence of hunger between 2001- 2019

**4.4.4 World Hunger Trends over the Years**

To begin the analysis, the data related to the entire world was extracted which contains information on various countries, has specific entries for "World" that aggregate data for the global prevalence of undernourishment.

With the world data selected, a line plot was created. In this plot, the x-axis represents the years and the y-axis represents the prevalence of hunger in percentage. Each point on the line graph corresponds to a specific year displaying how the global prevalence of hunger has changed over time. To make the plot visually appealing and aligned with a cyberpunk style, the mplcyberpunk library was used. This library adds a futuristic and cyberpunk-inspired glow effect to the plot, enhancing its visual impact. The line plot showed a decreasing trend from 2001 to 2018 and rise in trend line showing rise in world hunger in 2019.

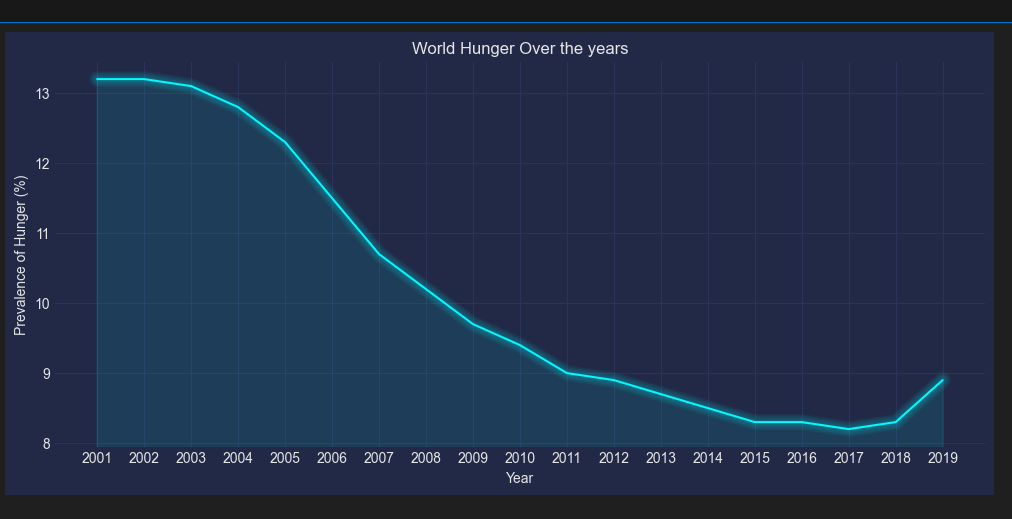
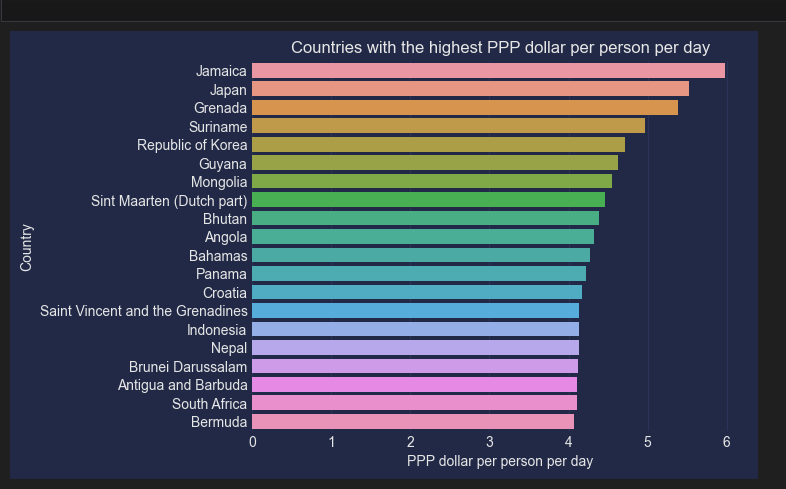


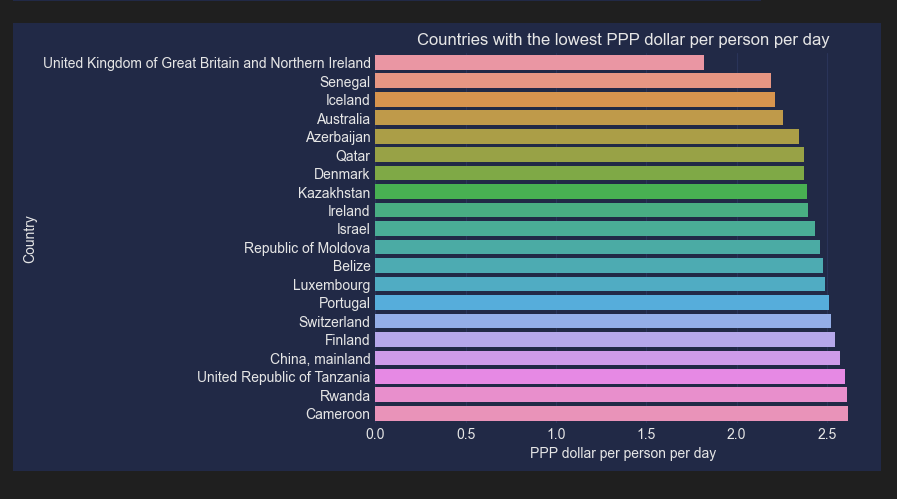
Fig 4.4.4: Line plot showing the world hunger over the years

**4.4.5 Analysis of the Costliest PPP Dollar per Person per Day**

Visualization

To identify and visualize the countries with the highest PPP dollar per person per day in 2017, a bar chart is created. This bar chart highlights the top 20 countries with the most expensive PPP values. The resulting bar chart provides a clear and concise visualization of the countries where individuals had the highest and lowest. The chart shows Jamiaca as being the highest with United Kingdom and Nothern Ireland as the lowest purchasing power per person per day.





**4.4.6 Forecasting Hunger Prevalence**

Using advanced time series forecasting techniques, future trends in hunger prevalence for all countries in the dataset was predicted. To achieve this, the ARIMA (AutoRegressive Integrated Moving Average) model, a powerful tool for analyzing and forecasting time series data was employed.

**4.4.7 Analyzing Employment Data**

This analysis was initiated by acquiring employment data from a CSV file. The data preprocessing steps include selecting rows based on indicator codes and gender. Afterward, the necessary columns were extracted, reset the index, and rename them appropriately. The resulting processed\_employment\_df contains valuable information about the number of employed individuals across various countries.

**Analyzing Employment Statistics**

Once the employment data is prepared, we proceed to compute the mean number of employed individuals per country. This analysis helps us identify countries with the highest employment rates. The bar chart shows that India has the highest employed indivduals and Marshal Island having the lowest.

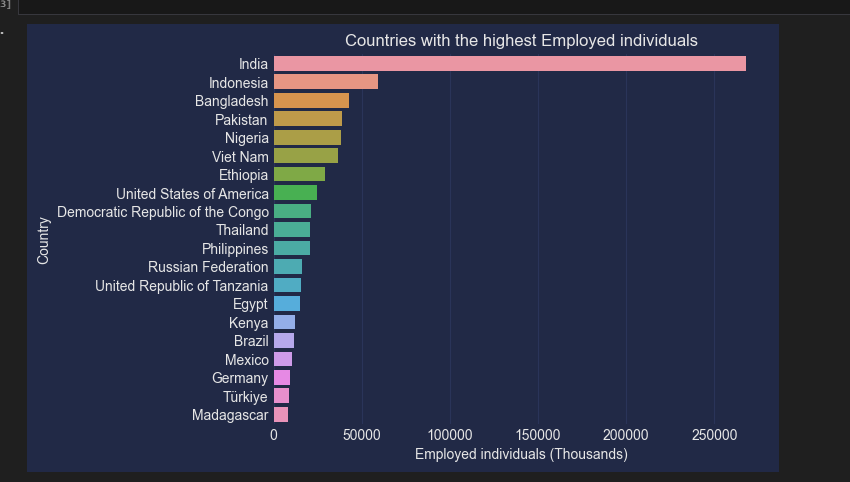


Fig 4.4.6 Countries with the highest employed individuals

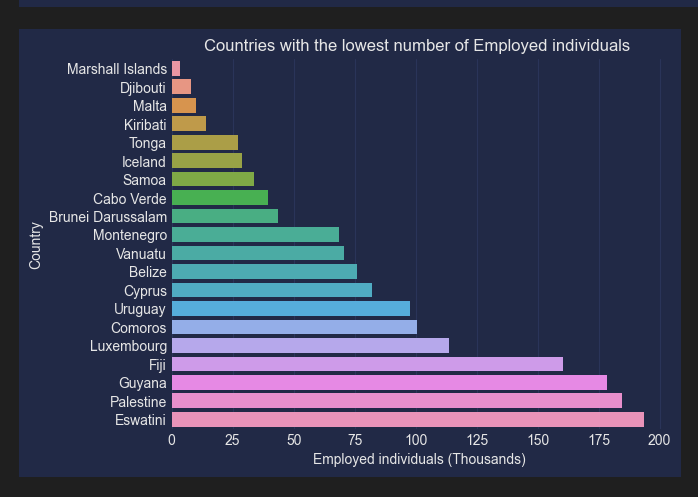


Fig 4.4.6 Countries with the lowest employed individuals

**4.4.7 Countries with the Highest and Lowest Food Price Indices (2001)**

Forecasting Food Inflation

Using the ARIMA model once again, food price inflation and consumer price indices for countries were forecast. This enables us to anticipate future trends in food prices and related indices. Lebanon has the highest food inflation while South Sudan has the lowest

**4.5 Hunger Prediction**

**4.5.1 Data Loading and Source Introduction**

The code begins by loading several datasets essential for the analysis. These datasets are sourced from various reputable sources, such as FAOSTAT, and contain crucial information related to hunger, food affordability, employment and food prices. This data amalgamation is crucial as it allows for a comprehensive analysis of hunger's relationship with other socioeconomic indicators.

**Data Preprocessing and Cleaning**

One of the initial steps in data preprocessing involves merging the datasets into a unified DataFrame. This step is accomplished using common columns, 'Country' and 'Year,' which act as key identifiers. Merging these datasets enables a holistic exploration of relationships between hunger and various economic indicators. Addressing missing data is the next critical step. It identifies and handles missing values efficiently, ensuring that the dataset is complete and ready for analysis. Any columns with missing data are imputed with the median value to maintain data integrity.

**Feature Engineering**

Feature engineering is essential for constructing meaningful predictive models. Lag features are generated for several economic indices. These lag features represent historical data for each index, providing valuable context for predicting hunger. For instance, 'PPP dollar per person per day\_prev1\_year' represents the value of PPP (Purchasing Power Parity) from the previous year for a given country.

**Data Normalization**

To ensure that the machine learning models operate effectively, continuous features are standardized using z-score normalization. This process scales these features to have similar ranges preventing one feature from dominating the modeling process due to its magnitude.

**One-Hot Encoding**

Categorical features, such as 'Country,' are transformed using one-hot encoding. This technique converts categorical variables into numerical format, making them suitable for machine learning algorithms. In this case, 'Country' is converted into a series of binary columns, one for each unique country in the dataset.

**Machine Learning Models**

Three distinct machine learning models were introduced: linear regression, decision tree regressor, and random forest regressor. These models are chosen to predict hunger prevalence based on the prepared dataset. They serve as tools for understanding the relationships between socioeconomic indices and hunger.

**Model Evaluation Metrics**

Several evaluation metrics were defined in order to assess the performance of these models,including Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-squared (R2) Score.

**Model Fitting and Evaluation**

The dataset was splitted into training and testing sets, a fundamental practice in machine learning. It uses 70% of the data for training and 30% for testing. The linear regression model is then fitted to the training data and evaluated using the defined metrics. The process is repeated for the decision tree regressor and random forest regressor.

**Visualization of Results**

To provide a clear understanding of the model performance, a visualization section was included. Bar charts are generated for each evaluation metric, displaying the performance of the different models. These visualizations help stakeholders and analysts interpret the model results effectively.

**Feature Importance Analysis**

In addition to model evaluation, a feature importance analysis for the Random Forest model was conducted. This analysis identifies which features (economic indices) have the most significant impact on predicting hunger prevalence. The top features are highlighted in a bar chart, shedding light on the key factors influencing hunger.

**4.5.2 Conclusion**

In conclusion, this is a comprehensive analysis of predicting hunger prevalence based on various socioeconomic indices. It demonstrates the power of machine learning in understanding complex relationships between economic factors and hunger. The results and visualizations generated offer valuable insights for policymakers and organizations working to combat food insecurity.

**4.6 Visualization Setup**

**4.6.1 Introduction**

Essential libraries were imported including pandas, numpy, seaborn and matplotlib. These libraries are commonly used for data manipulation and visualization. Additionally, it installs the mplcyberpunk library which is used to give the plots a cyberpunk-style appearance. It also suppresses warnings for a cleaner output.

Next, the code loads data from an Excel file named 'Hunger and Undernourishment Data-1.xlsx' into a pandas DataFrame called 'df.' This dataset likely contains information related to hunger and undernourishment. Additionally, the code performs a minor data transformation replacing country names to ensure consistency.

**4.6.2 Geospatial Data Handling**

The code introduces geospatial visualization by utilizing the geopandas and folium libraries. It loads a GeoJSON file named 'world\_countries.json,' which likely contains geospatial data for countries and regions. A base world map, using the 'cartodbpositron' tileset, is initialized with an initial view centered at latitude 0 and longitude 0.

**4.6.3 Choropleth Map Generation**

A choropleth map is generated using the folium.Choropleth function. This map visualizes the prevalence of undernourishment across different countries. It is based on the 'df' DataFrame and uses the GeoJSON data for country boundaries. The color of each country is determined by the 'Prevalence of undernourishment' column, with a color scale from yellow to red representing low to high prevalence. A legend is included for reference.

**4.6.4 Data Integration with Choropleth**

To display data on the choropleth map, the code performs data integration. It associates country-specific information, such as the 'Prevalence of undernourishment,' with the respective countries on the map. This is achieved by matching the country names in the GeoJSON data with those in the 'df' DataFrame.

**4.6.5 Tooltip Addition**

To enhance user interaction with the choropleth map, tooltips are added. These tooltips display information about each country when the user hovers over it. The information includes the country name and its 'Prevalence of undernourishment (%)' value.

**4.6.6 Layer Control**

The code adds a layer control to the map, allowing users to toggle between different layers or views. This control typically enables the user to switch between different datasets or map features.

**4.6.7 Country Code Retrieval**

A section of the code retrieves the country codes (short codes) for each country in the DataFrame. These codes are often used for various geographical and analytical purposes. Any country name inconsistencies are addressed in this step.

**4.6.8 Data Exploration**

The code includes exploratory data analysis (EDA) by visualizing the prevalence of hunger using a pygal world map. This map represents the 'Prevalence of Hunger (%)' in the year 2005. Data for this visualization is extracted from the 'df' DataFrame.

**4.6.9 Rendering and Output**

The code concludes by rendering the generated world map with the 'Prevalence of Hunger (%)' data for 2005 as an SVG file named 'world\_map2.svg.' This file likely contains the final visual representation of the hunger prevalence worldwide in 2005.

Additionally, there are some commented-out sections that appear to be related to creating choropleth maps for other data, such as 'PPP dollar per person per day.' These sections follow a similar structure to the 'Prevalence of undernourishment' choropleth generation but are commented out, suggesting they are not currently active in the code.

**CHAPTER FIVE**

**RESULTS**

**5.1 Introduction**

The results of three different machine learning models were employed to predict hunger prevalence. The models under consideration are Linear Regression, Decision Tree Regressor and Random Forest Regressor. The aim is to assess the predictive performance of them using various evaluation metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and R-squared (R2) score. Beyond just presenting the metrics, the significance and implications of these results was explored shedding light on how these models can contribute to understanding and addressing critical global challenges related to hunger.

**5.2 Model Evaluation Results**

**5.2.1 Linear Regression**

The Linear Regression model emerges as the star performer in this analysis. With an R-squared (R2) score of 0.91, it demonstrates an exceptional ability to predict hunger prevalence. The R2 score which measures how well the model explains the variance in the data, is close to its maximum value of 1. This indicates that the Linear Regression model captures nearly all the underlying patterns and relationships within the dataset. Furthermore, the low Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) values (12.17, 3.49 and 2.23, respectively) underline the model's accuracy in approximating actual hunger prevalence rates.

**5.2.2 Decision Tree Regressor**

The Decision Tree Regressor while not matching the accuracy of the Linear Regression model, provides valuable insights into hunger prediction. Its R2 score of 0.58 indicates a moderate ability to explain variance, suggesting that it captures some but not all patterns in the data. The higher error metrics including an MSE of 54.61, RMSE of 7.39, and MAE of 5.33 imply that this model has a wider margin of error compared to Linear Regression. Despite its limitations, the Decision Tree model's simplicity and interpretability make it a valuable tool for understanding the key factors affecting hunger prevalence.

**5.2.3 Random Forest Regressor**

The Random Forest Regressor offers a middle ground between the precision of Linear Regression and the interpretability of the Decision Tree. With an R2 score of 0.70 which demonstrates a good fit to the data, explaining a substantial portion of the variance. The error metrics, with an MSE of 38.23, RMSE of 6.18 and MAE of 4.80 fall between the other two models. This indicates that Random Forest strikes a balance between accuracy and complexity, making it a practical choice for predicting hunger prevalence in real-world scenarios.

Additionally, the feature importance analysis of the Random Forest model offers insights into the key factors affecting hunger prevalence. Decision-makers can use this information to prioritize interventions and allocate resources effectively. For instance, if employment rates and food prices emerge as critical factors, policies aimed at improving employment opportunities and stabilizing food prices can be emphasized.



Fig 5.2: A bar chart showing the mode’s results

**5.3 Overall Insights and Implications**

The results of these machine learning models have several important implications for addressing hunger-related challenges:

* **Understanding Causality:** The Linear Regression model's strong performance suggests that certain factors have a direct and linear impact on hunger prevalence. Identifying these causal factors can guide targeted interventions.
* **Non-Linearity Matters:** The Decision Tree and Random Forest models emphasize the presence of non-linear relationships in the data. Policymakers must consider both linear and non-linear factors when formulating strategies
* **Resource Allocation:** Decision-makers can utilize the insights from these models to allocate resources effectively. By prioritizing factors with high feature importance, interventions can be tailored for maximum impact.
* **Model Refinement:** The results also point towards opportunities for further model refinement. Hyperparameter tuning and more sophisticated ensemble techniques could potentially enhance prediction accuracy.

In conclusion, these machine learning models provide a data-driven approach to understanding hunger prevalence.

**CHAPTER SIX**

**CONCLUSION**

The analysis and prediction of hunger prevalence through machine learning models have yielded profound insights into the multifaceted factors influencing hunger and undernourishment on a global scale. This comprehensive study integrated diverse datasets encompassing economic indicators, affordability indices, employment statistics and food price trends to construct predictive models. The implications of these models offer a data-driven foundation for comprehending and addressing critical challenges related to hunger and undernourishment worldwide.

These findings have profound implications for policy formulation and international efforts to combat hunger. Firstly, the strong performance of the Linear Regression model suggests that addressing hunger requires a focus on economic growth, job creation and improving food affordability. Policies aimed at boosting GDP, promoting employment, and implementing food subsidy programs can play a crucial role in reducing hunger.

Secondly, recognizing the existence of non-linear relationships as indicated by the Decision Tree model, emphasizes the need for holistic and context-specific solutions. Policymakers must consider a wide range of factors, from cultural practices to climate change when designing effective interventions.

Lastly, the insights gained from machine learning models can enable the design of precisely targeted interventions. By prioritizing factors with high feature importance, governments and organizations can allocate resources more efficiently to regions and populations most in need maximizing the impact of their efforts.

It is essential to acknowledge the limitations of this analysis. The predictive models rely on historical data. Additionally, data availability and quality can significantly affect model performance. Future research can explore the integration of real-time data and validation procedures to enhance the robustness of predictions.

Summarily, the analysis and prediction of hunger prevalence offer a data-driven framework for addressing one of the most pressing global challenges. These models provide valuable tools for understanding the intricate nature of hunger and undernourishment by empowering policymakers and organizations to take evidence-based actions aimed at reducing hunger, enhancing food security, and ultimately improving the well-being of communities worldwide.