**Environmental impacts of food production and consumption**

Vaishnavi Lalwala, Sargam Malaviya, Dhvani Chokshi

Computer Engineering, School of Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat

**Abstract:**

This comprehensive review explores the intersection of data analytics and sustainable food production by synthesizing findings from a range of research papers. It investigates methodologies for assessing the environmental impacts of food production and consumption, including life cycle assessment (LCA), input-output analysis, and empirical studies. Additionally, it delves into the application of machine learning techniques in addressing Sustainable Development Goals (SDGs) related to food security, sustainable agriculture, and climate action. The review scrutinizes various modeling approaches employed to predict the environmental consequences of agricultural practices, highlighting their strengths and limitations. Furthermore, it discusses machine learning applications in environmental monitoring, emphasizing their role in advancing our understanding of ecosystem dynamics and informing sustainable management practices. By synthesizing insights from diverse research streams, this review underscores the potential of data analytics in shaping future environmental policies and practices to promote sustainable food production.

**Introduction:**

This project therefore aims at providing a general **overview** of the **global food supply** of different food items and different countries and to analyse the **environmental impact** of different food products.

**Dataset Description:**

1. **FAO 'Food Balance Sheets' Dataset:**
   * Source: Food and Agriculture Organization of the United Nations (FAO)
   * Description: This dataset provides comprehensive information on the food supply patterns of 174 countries spanning from 1961 to the latest update available in 2013. The data is measured in 1000 tonnes units, offering insights into the quantity of various food items available for consumption in each country over the specified time period.
2. **'Environment Impact of Food Production' Dataset:**
   * Description: This dataset focuses on the environmental impact of food production, encompassing 43 commonly consumed foods grown globally. It comprises 23 columns representing various environmental metrics such as land usage, water consumption, and carbon footprints associated with the production of each food item. These metrics are expressed in kilograms of CO2 equivalents per kilogram of food product, covering different stages in the lifecycle of food production.

**Literature Survey:**

Life Cycle Assessment (LCA) is pivotal for quantifying environmental impacts across product and service life cycles, particularly in comparing open and indoor farming practices due to its comprehensive approach. Assessment often relies on key indicators like water, carbon, and ecological footprints, offering insights into resource use and environmental burdens. Computational methods like Artificial Neural Networks (ANN) are increasingly applied, demonstrating efficacy in optimizing agricultural processes such as methane production and crop yield prediction. Addressing data challenges, researchers utilize various imputation methods and feature compression for complex datasets. Visualization techniques including catplots, heatmaps, and boxplots aid in communicating nuanced insights, such as product-specific emissions and operational impacts across the value chain.

**Project Implementation:**

1. Data Exploration, Cleaning, and Transformation:
   * Begin with a standard approach to explore the dataset, examining variable names, types, content, and overall structure to gain a comprehensive understanding of the data.
   * Clean the data by removing unnecessary rows and columns, handling missing data, and eliminating duplicates.
   * Transform the 'Food Balance Sheets' dataset for improved visualization and data handling, addressing negative values where necessary.
2. Handling Missing Data: Imputation Methods:
   * Identify the "Missing Not At Random" (MNAR) archetype in the dataset, indicating a systematic bias in the missing pattern.
   * Evaluate various imputation methods including dropping columns with missing values, imputation by average value, median, K nearest neighbors (KNN), multiple imputation by chained equations (MICE), and custom-fitted imputation methods.
   * Choose three columns with the highest amount of missing values as examples to illustrate the results of each imputation method and compare the differences in mean values after each method.
3. Visualization and Analysis:
   * Set up visualization tools and frameworks.
   * Implement MICE imputation using Bayesian Ridge as the chosen method for dealing with missing values.
   * Generate a useful overview visualization depicting Total Emissions by each Food Product, filtered by food Category, to provide insights into emission patterns across different food categories.
   * Create tree maps to visually represent the distribution of emissions by product category.

**Expected Outputs:**

1. **Top of List for Total Emissions:**
   * Identify the food products with the highest total emissions, providing insight into the primary contributors to environmental impact within the dataset.
2. **Useful Overview Visualization of Total Emissions by Food Product and Category:**
   * Create a visualization showcasing total emissions by each food product, filtered by food category. This visualization offers a comprehensive understanding of emission patterns across different food categories.
3. **Food Supply vs. Population Increase:**
   * Analyze the relationship between food supply and population increase over time, shedding light on the dynamics of food production and consumption in relation to population growth.
4. **Impact of Water Usage on Emissions:**
   * Investigate the influence of water usage on emissions, exploring the correlation between water consumption and environmental impact within the context of food production.
5. **Correlation Analysis of Variables Influencing Total Emissions:**
   * Utilize a "coolwarm" heatmap and pairplot to visually depict the correlation between total emissions and other relevant variables. This analysis aims to identify key factors influencing emissions within the dataset.
6. **Emissions by Product Category:**
   * Examine emissions distribution across different product categories, providing insights into the relative environmental impact of various food categories.
7. **Agriculture Operational Emissions by Value Chain:**
   * Assess operational emissions along the agriculture value chain, highlighting emissions associated with different stages of food production, processing, and distribution.
8. **Creation and Visualization of a Normalized Variable Index:**
   * Develop a normalized variable index to condense complex emission data into a meaningful format. By aggregating multiple emission-related variables, this index facilitates a concise overview of environmental impact for both individual food products and categories. Visualizations of the index enable stakeholders to glean insightful summaries without overwhelming them with numerous graphics.

**References:**

# Estimating energy consumption and GHG emissions in crop production: A machine learning approach.

Saeed Sharafi a b, Ali Kazemi a b, Zahra Amiri .

<https://www.sciencedirect.com/science/article/abs/pii/S0959652623014002#preview-section-references>.

# Using the Machine Learning Method to Study the Environmental Footprints Embodied in Chinese Diet

# by [Yi Liang](javascript:void(0))1,[Aixi Han](javascript:void(0))2,[Li Chai](javascript:void(0)) and [Hong Zhi](javascript:void(0))

# [Life Cycle Assessment](https://www.sciencedirect.com/science/article/pii/B9780123864543006278)- G. Finnveden, J. Potting, in [Encyclopedia of Toxicology (Third Edition)](https://www.sciencedirect.com/referencework/9780123864550/encyclopedia-of-toxicology), 2014

1. Environmental impacts associated with food production

2021 | Dr Martin Juneau, M.D., FRCP

1. Food production is responsible for one-quarter of the world’s greenhouse gas emissions By: [Hannah Ritchie](https://ourworldindata.org/team) *November 6, 2019*
2. Analysis, Visualization, and Extensive Imputation of CO2 Emissions in Food Products.