Midterm Project

Impact of Food Production on Environment

Team Machine Mavericks

Devi Priya Ravi
Dhanasree Rajamani
Sravani Thota
Abdul Vahed Shaik

Objective

The objective of the project is to achieve food sustainability by identifying the sources of greenhouse gas emissions in food production. This includes determining the food products which are significant contributors to greenhouse emissions, the stages of production with the highest emissions, and the countries that produce these food products. Through this analysis, we can determine the role of food producers and consumers in promoting food sustainability. For this purpose we are using the following datasets:

[Environmental Impacts of Food Production](#)

[Food and Agriculture data](#)

Source : [Our World in Data](#)

Goal Service Model

Goals

The world's growing population is driving an immense demand for food production, without equal attention for its environmental impact. Animal-based food products consume more resources and contribute to a larger carbon footprint compared to plant-based diets. It is highly essential that we are cautious regarding our food choices and make a conscious shift towards more viable eating habits. This considerate change in the living can bring in a positive change in the long run of one's being towards a better future and good health.

The goal of the project is to develop a model that can predict the likelihood of countries adopting sustainable production techniques based on factors such as

- Measuring and setting threshold value on emissions
- Focus on alternative sources of energy
- Promote and bring in more plant-based diets
- Reduce Food waste

Additionally, the model could be used to evaluate the effectiveness of the approach for building sustainable diets, and to develop recommendations for the most effective strategies.

Metrics – measured by Accuracy, Loss, Confusion matrix, Precision, Recall, RMSE, R2 values, AUC (Area Under ROC curve), Mean Absolute Error (MAE), Root Mean Square Error (RMSE)

- Data

    We have chosen 2 datasets now:

    Environmental Impacts of Food Production

    Food and Agriculture data

Refinements planned to perform : Data cleaning, Data imputation, Data Preprocessing, Dimensionality reduction, Feature engineering

- Application / microservices

Applications include:

- Sustainable agriculture
- Crop yield prediction
- Energy efficiency

- U/X

    Tableau/ Power BI dashboard

- Business Logic :The business logic is directed towards improving the sustainability of agricultural products and hence reducing the environmental impact of food and feed production. This leads to developing the performance and benefit of the agriculture sector. By investing in ML technology, companies in the agriculture field could lead in sustainable food and feed production, potentially alluring more environmentally-conscious samaritans and investors.

- Models :
    - Clustering(K means, Agglomerative)
    - Regression (Linear regression, Gradient boost).
    - Classification (Decision tree, Random forest, Support Vector Machine, Naive bayes).

Determine the best model for the data based on various measures like R2 score, RMSE.

Data Narrative

1. What is the data telling me?

The datasets provide information on:

- **Impact of Food production on environment**

    This dataset consists of most common(popular) foods grown around the world along with their land usage, eutrophication, water usage, and carbon footprints.

- **Production and consumption of food**
    This dataset consists of the most common(popular) foods grown around the world - with the countries in which they are grown, quantities in which they are produced each year(1961-2020) and who consumes it(human, livestock, poultry)

2. What can the data tell me?

From this dataset, we can answer the following questions:
- Which food item is grown the most?
- Which food item emits the most and least greenhouse gasses?
- Which countries are leading producers of different food items?
- Which stage of the food production(land usage, transport, etc.) emits most greenhouse gasses?

3. Patterns you see in the data

Regional trends: The data set can be analyzed for identifying the regional trends in food and feed production and its effect on the surroundings, such as differences in land use, water use, greenhouse gas emissions, and chemical use between regions.

Environmental impact per unit of food and feed: The data set can be reviewed to determine the environmental impact per unit of food and feed produced, such as the amount of greenhouse gas emissions, water use, or chemical use per kilogram of food or feed.

4. What are you looking for in the data

Dairy vs. plant-based milk: what are the environmental impacts?
Yields vs. Land Use: How the Green Revolution enabled us to feed a growing population.
How do food and feed production and its impact on the environment vary by region or country?
Which country has most and least sustainable food production methods?

Topics to consider

https://towardsdatascience.com/are-you-still-using-the-elbow-method-5d271b3063bd

https://www.cambridge.org/core/journals/data-and-policy/article/food-security-analysis-and-forecasting-a-machine-learning-case-study-in-southern-malawi/CA4DFA39526F318373259921C10D1C3F

https://www.nature.com/articles/s43016-022-00587-8

https://towardsdatascience.com/how-to-mesure-the-carbon-footprint-using-vertex-ai-pipelines-3d6bc9695e7b

https://medium.com/codefile/data-analysis-with-chatgpt-and-jupyter-notebooks-fa2b03753396

https://medium.com/mlearning-ai/how-data-science-is-evolving-the-food-industry-1ad02f3f9952