A Data-Driven Analysis of Agricultural Sustainability: Insights from Emissions, Land-Use Efficiency, and Sustainability Prediction for Top Agricultural Producers

ID - U5670402

Abstract - *This study focuses on agricultural sustainability in the top eight global producers, namely China, India, the United States, Brazil, Indonesia, Russia, Mexico, and Japan. Using two primary datasets, we look at CO_2 emissions, land-use efficiency, and sustainability through exploratory data analysis and predictive modelling. The Logistic Regression made accuracy of sustainability levels classification as 92%, and Gradient Boosting with SHAP revealed the key emission drivers, such as industrial processes and food consumption. The findings stress the need for precision agriculture, cleaner industrial practices, and energy efficient policies to enhance sustainability. This work is a contribution to the global discussions on food security and environmental protection.*

I. Introduction

Agriculture is one of the most important sectors for humans and global economy as well, it helps us in achieving food security which is one of the threats to us humans in the coming future with climate change still on a rise. While it does support economic development as well but, at the same time, it is one of the major causes of greenhouse gas (GHG) emissions and puts pressure on natural resources. Emissions from livestock, rice cultivation, fertilizer use, and deforestation bring CH_4, N_2O and CO_2 emissions which enhance climate change. These environmental impacts must be addressed to develop a more sustainable and resilient agricultural system. (Smith et al., 2014).

The top eight agricultural producers: China, India, United States, Brazil, Indonesia, Russia, Mexico and Japan face a major challenge to reduce their emissions due to agriculture. These nations generate a large share of the world's food while also being the major emitters of agricultural emissions. They need to work on enhancing agricultural production to meet the increasing demand while reducing GHG emissions. The barriers to sustainability are substantial because of problems such as inefficient land use, high resource consumption and inadequate emissions reduction strategies. (Tilman et al., 2011).

These challenges can be tackled using data driven approaches. Analysing emissions and land use data enables us to determine patterns, where there is scope for improved sustainability, and what actions can be taken. This study uses exploratory data analysis (EDA) and predictive modelling to examine the key trends and drivers of agricultural sustainability across the top eight producers. Derived metrics such as emissions intensity and land use efficiency are used to compare these nations and improve understanding of areas for improvement.

We use Logistic Regression for classifications sustainability levels and use Gradient Boosting models to identify the most influential factors driving emissions and inefficiencies. Through these methods we can gain a deeper understanding of the factors that influence sustainability outcomes and guide data-informed decision making in agriculture.

Contributing meaningful insights to the global sustainability dialogue, this report focuses on the world's largest agricultural producers. Findings from our report are intended to inform actionable policies that promote sustainable agricultural practices and ensure food security as well as environmental protection.

II. Data Overview

This study uses two primary datasets to analyse CO_2 emissions, land use, and sustainability across the top agricultural producers globally.

The **Agrofood CO_2 Emission** dataset (Lobello, n.d.) offers a holistic perspective of the CO_2 emissions that result from several agricultural activities. Emissions from rice cultivation, livestock and food processing are included in the data while also including the emissions absorption by forests. In addition to emissions data, the dataset contains rural and urban population distributions, and average annual temperature increase, among other demographic and environmental indicators. This dataset serves as the initial platform for the analysis of the environmental effects of agricultural practices.

The **Agricultural Land Area dataset** (World Bank, n.d.) adds crucial context by detailing the total land area allocated to agricultural use for various countries over time. We calculated the agricultural land use with our emissions dataset to measure the countries efficiency in utilizing the resources and balancing emissions for environmental sustainability.

We combined these datasets on Country and Year, gathering everything into a single framework to analyse. From this integration, we were able to derive key metrics, including:

- **Emissions Intensity**: CO_2 emissions per capita, which consider the environmental footprint in relation to population, providing a perspective on the societal impact of emissions.
- **Land-Use Efficiency**: CO_2 emissions from agricultural land: a bottom-up perspective on the environmental costs of practices due to Land Utilization efficiency.
- **Agricultural Emissions Share**: The share of total CO_2 emissions from agriculture, measuring the sector's emission contribution to total emissions.

Our analysis is focused on the top eight agricultural producers: China, India, the United States, Brazil, Indonesia, Russia, Mexico and Japan. These countries are pivotal in food production and agricultural emissions on a global level and are therefore key to understanding and addressing sustainability challenges.

III. Methodology

a. Data Preprocessing

We began by preparing the datasets to ensure consistency and reliability for analysis.

Transforming the Agricultural Land Area Dataset
First, we normalized the Agricultural Land Area dataset so that it is consistent with the structure of the Agrofood CO_2 Emission dataset. We merged the two datasets using 'Area'(Country) and 'Year' as common keys to create a unified framework. This transformation was useful to be able to analyse emissions and land use together for a comprehensive sustainability metrics view across countries.

Handling Missing Values
To handle missing values I used linear interpolation for numerical features, so that the temporal trends within each country are preserved. I removed incomplete rows for categorical values to maintain data integrity. These steps ensured that the dataset was robust and suitable for further analysis.

Feature Engineering
We created several key metrics to improve our analysis:

- **Emissions Intensity**: This metric is used to normalize CO_2 emissions by population size to understand the per capita environmental impact.
- **Land-Use Efficiency**: We have calculated emissions per unit of agricultural land to reveal the environmental cost of land use.
- **Agricultural Emissions Share**: This metric is used to express the share of total CO_2 emissions due to agriculture to find out the contribution of the sector to the total emissions.

They are the engineered features for our exploratory and predictive analyses.

b. Exploratory Data Analysis (EDA)

We used a detailed EDA to find patterns, relationships and trends in the data.

Visualizing Key Metrics

We used temporal trend plots to show the changes in emissions intensity, land-use efficiency and agricultural emissions share over time for the top 8 producers. For example:

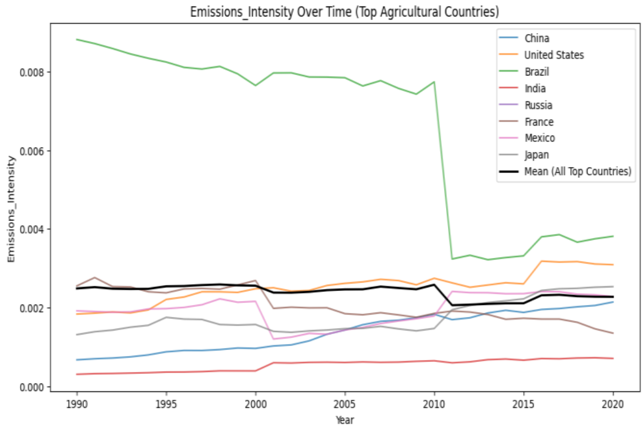

Fig. 1 Emissions intensity over the period of 30 years from 1990 – 2020 for top agricultural producers

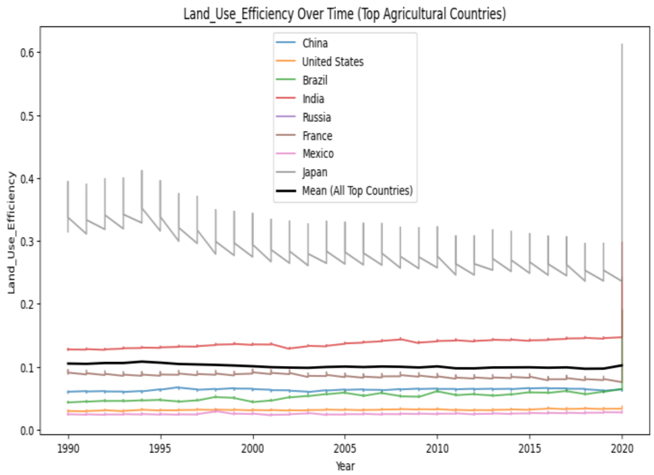

Fig. 2 Agricultural land use efficiency over the period of 30 years from 1990 – 2020 for top agricultural producers

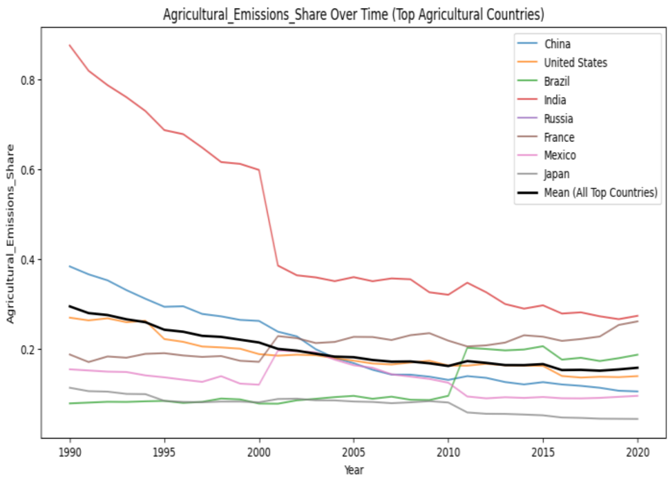

Fig.3 Agricultural Emissions over the period of 30 years from 1990 – 2020 for top agricultural producers

- From Figure 1, we see that emissions intensity tends to decrease in most countries, with some differences still existing.
- Figure 2 shows land-use efficiency trends, with some nations improving significantly while others stagnate.
- The share of agricultural emissions is captured in Figure 3, which shows the changing emissions from agriculture.

Understanding Feature Relationships

Through correlation heatmaps, we were able to see the relationships between variables and use that to choose the best variables to put into our predictive models. These insights informed the structure and focus of the next phase of the analysis.

c. Predictive modelling

We used predictive modelling to classify sustainability levels and to identify the most significant drivers of emissions and inefficiencies.

Logistic Regression for Sustainability Classification

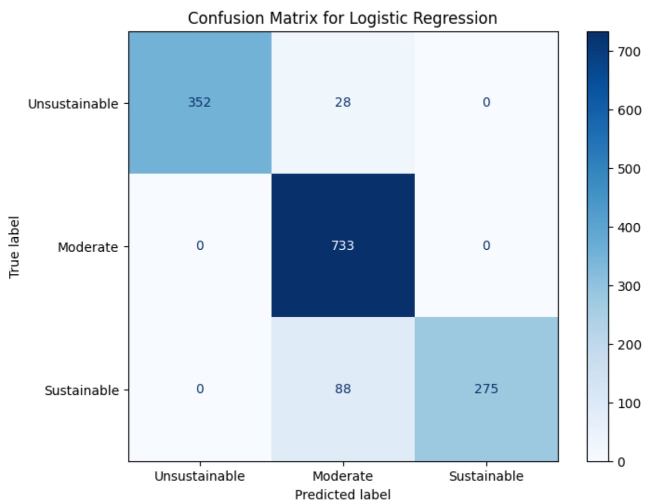

Fig.4 Confusion matrix for logistic Regression

Using derived metrics such as emissions intensity and land use efficiency as predictors, we applied logistic regression to categorize countries into three sustainability levels: "Sustainable," "Moderate," and "Unsustainable." The logistic regression model showed robust classification performance, with a high accuracy of 92%, as demonstrated by the confusion matrix in Figure 4.

Gradient Boosting for Feature Importance

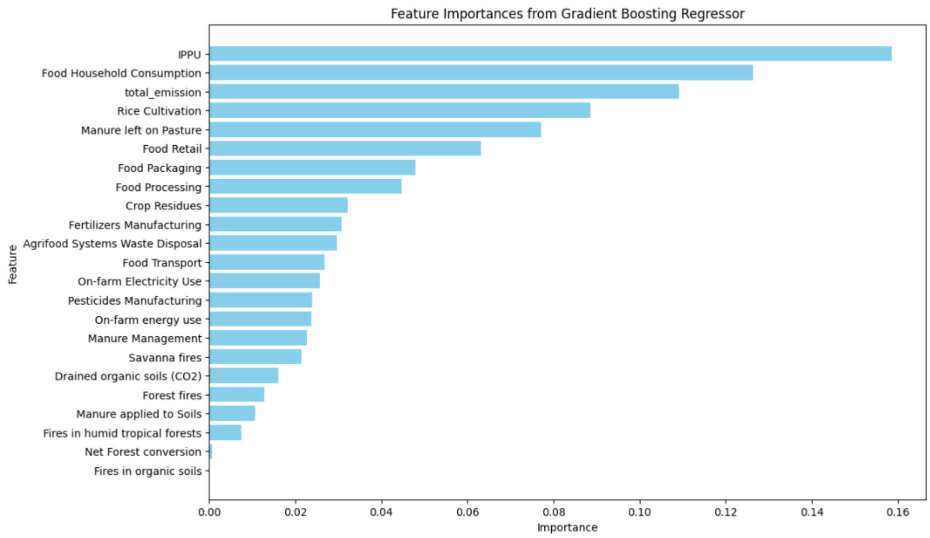

Fig.5 Feature importance from our gradient boosting

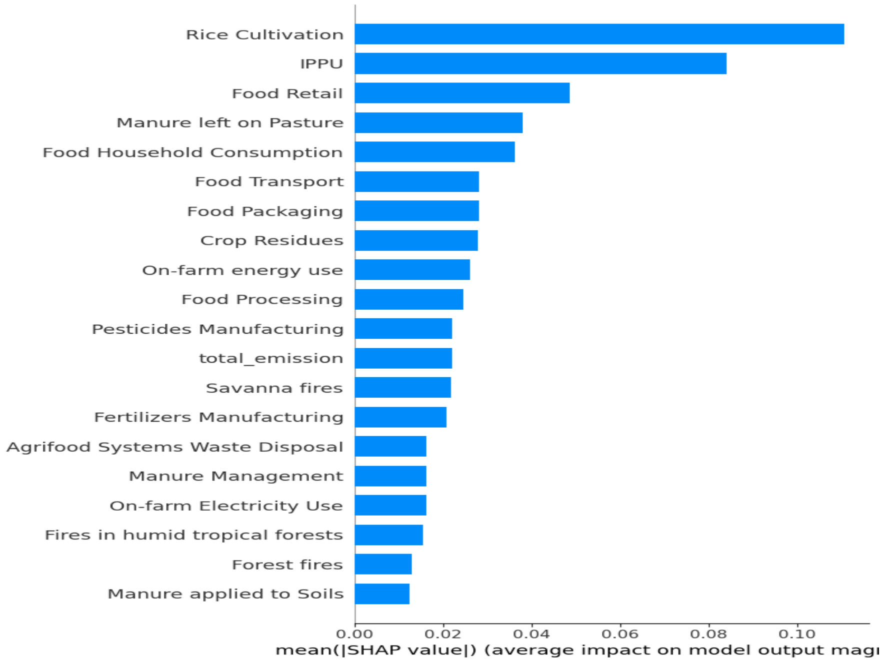

Fig.6 Feature importance from our gradient boosting using SHAP

To understand the factors driving sustainability, we used a Gradient Boosting Regressor combined with SHAP (SHapley Additive Explanations). Using this approach, the most influential features that contribute to emissions and inefficiencies were revealed. Some of the key findings are:

- From the Figure 5 emissions breakdown, it appears that 'Industrial Processes and Product Use (IPPU)" and "Food Household Consumption" are the biggest contributors to emissions, and hence they are most critical.

- Through SHAP analysis in Figure 6 we get additional feature interactions insights like how Rice Cultivation and Manure Management are indirectly impacting sustainability outcomes.

The Gradient Boosting model performed exceptionally well, achieving an R^2 score of 0.9855 and an MSE of 0.0031. These results confirmed the reliability of our feature selection and model performance.

Using rigorous data preprocessing, insightful EDA, and advanced predictive modelling, we identified critical drivers of agricultural emissions and sustainability when combining them. This methodology not only provides a comprehensive understanding of the environmental challenges faced by top agricultural producers but also lays the groundwork for actionable policy recommendations.

IV. Results and Discussion

From our exploratory data analysis (EDA) and predictive modelling, this section presents findings on emissions, land-use efficiency, and sustainability levels across the top eight agricultural producers. We offer a comprehensive perspective on agricultural sustainability by comparing our results with existing research.

Through emissions intensity analysis, we found that emissions have generally decreased in most countries. Let's look at the case of Brazil, which has made noticeable progress, which may be due to better agricultural practices and emissions control measures. In contrast, India and Indonesia had relatively weak reductions in emissions, which indicate that these countries still have challenges to overcome to improve the efficiency of their agriculture. These findings are consistent with those of Yasmeen et al. (2022), who found large variations in carbon emission efficiencies among developing countries. The trends observed further stress the importance of targeted measures in high emission areas to improve agricultural systems.

The efficiency of land use was far from uniform among the top producers. The United States and Brazil showed steady increases, which reflect the implementation of new, more efficient farming methods and better land management. But China's efficiency has been stagnant, implying the lack of new ideas in land use. Similarly, Zhang et al. (2023) also noted that enhancing land-use efficiency is vital for sustainable agricultural growth and development. Thus, the emphasis is rightly placed on precision agriculture and policies that support sustainable land use.

The analysis of the share of agricultural emissions showed different trends. The emissions share in India has been steadily decreasing, which may be due to the growing use of renewable energy and the transition to green modern farming. By contrast, countries such as Brazil and Indonesia kept relatively constant shares, which indicates that there has been little progress in reducing deforestation and methane from livestock. Hayek et al. (2020) stresses on the topic of carbon opportunity cost of agricultural land, especially in areas with land-use change emissions, supporting the need for policies aimed at these regions.

Using our Logistic Regression model, we were able to effectively categorize countries into three sustainability levels: "Sustainable," "Moderate," and "Unsustainable" with an accuracy of 92%. Emissions intensity and land use efficiency were identified as key predictors by the model. For instance, countries labelled as "Sustainable" had low emissions intensity and high land use efficiency, whereas the "Unsustainable" countries had problems with inefficiencies in the use of resources. This is similar to other studies, including those of Zhang et al. (2023), which follow methodologies that focus on emissions efficiency as a primary indicator of sustainability.

Using Gradient Boosting Regressor with SHAP we were able to get a better understanding of what is causing emissions. The biggest contribution came from "Industrial Processes and Product Use (IPPU)" followed by "Food Household Consumption". These findings establish the downstream environmental consequences of industrial activities and consumer actions. Also, from the SHAP analysis we can see how there are interactions between features, for example, how "Rice Cultivation" increases emissions when paired with "Manure Management". These factors, including industrial processes and product use and food household consumption, which Yasmeen et al. (2022) identify as emitting greenhouse gases from agriculture, also suggest the need for mitigation strategies.

Our findings have important implications for sustainability. China and India which have stagnating or declining land–use efficiency must adopt technologies like precision agriculture to optimize their resource use and reduce emissions. Activities such as industrial processing and household consumption emissions require policy intervention encouraging the use of cleaner technologies and energy saving practices by giving incentives for its adoption. These problems can be fixed without great effort and would help to improve sustainability outcomes also contribute to global climate goals. These recommendations are consistent with the global calls for sustainable intensification of agriculture made by Tilman et al. (2011).

Through integrating our results with existing research this analysis gives a comprehensive view of agricultural sustainability among the top producers. Challenges and opportunities are highlighted by the findings, and actionable insights are offered for driving sustainable practices in agriculture.

V. Conclusion and Recommendations

This study provides valuable insights into the sustainability of agricultural practices among the world's top eight producers: China, India, the United States, Brazil, Indonesia, Russia, Mexico and Japan. Using exploratory data analysis (EDA) and predictive modelling, we identified key drivers of emissions, assessed land-use efficiency, and classified countries by sustainability levels. These findings reveal both progress and persistent challenges towards achieving sustainable agriculture.

Through our analysis, we found that there is a general decrease in emissions intensity for most countries, which is due to improvements in agricultural practices. However, there are still disparities; for example, India and Indonesia are making slower progress. These countries experienced mixed trends in land-use efficiency; the United States and Brazil made considerable gains, while China and India had little progress. These findings justify the requirement for targeted strategies tailored to each country's unique challenges.

Using the Logistic Regression model, emissions intensity and land use efficiency were identified as key predictors for sustainability levels. While the 'Sustainable' nations were found to have used their resources efficiently and had low emissions, the 'Unsustainable' nations had high emissions in relation to their land use. From the Gradient Boosting Regressor, "Industrial Processes and Product Use (IPPU)" and "Food Household Consumption" were found to be the most significant emitters. These findings are consistent with existing research in terms of the downstream sustainability impact of industrial and consumer activities.

Recommendations

1. **Adopt Precision Agriculture**
   China and India, and other countries, should devote their attention to adopting precision agriculture technologies to enhance land use efficiency and decrease emissions. Techniques such as site-specific nutrient management and optimized irrigation can greatly increase productivity with relatively weak impacts on the environment. Artificial intelligence and machine learning have advanced to the point that they have made precision agriculture more available and have revolutionized soil carbon measurement and resource management (Reuters, 2024a).

2. **Incentivize Cleaner Industrial Processes**
   Governments should back the adoption of cleaner technologies in agricultural processing and product use. Incentives such as subsidies or tax breaks can push industries to cut emissions from food manufacturing and packaging. Fertilizer production innovations such as green ammonia solutions that use renewable energy are already emerging as ways to decrease the carbon footprint of agricultural inputs (Reuters, 2024b).

3. **Promote Energy-Efficient Practices in Households**
   Household food consumption is also emitting a lot of emissions. Through educational campaigns and financial incentives, governments should encourage energy efficient appliances and reduce food waste. Sustainable consumption habits can be driven in a very concrete way from the consumer level.

4. **Strengthen Land-Use Policies**
   Countries with stagnating land-use efficiency should implement stricter land-use policies to prevent overexploitation of the agricultural land. Agroforestry and reduced tillage are two sustainable practices which can restore the soil health and enhance carbon sequestration. This is because addressing soil degradation is crucial, widespread degradation has been a cause for concern to global conservation organizations (Le Monde, 2024).

5. **Target High-Impact Activities**
   Rice cultivation and manure management are major sources of methane emissions. These challenges can be addressed effectively through promoting the System of Rice Intensification (SRI) and methane capture technologies for livestock. Research has also emphasized the urgency of prioritizing innovative practices for sustainable rice systems up to 2050 (Cambridge University Press, 2024).

Conclusion

Sustainable agriculture is the balance between food production and environmental harm. This study identifies key areas for improvement among the top agricultural producers and offers practical recommendations for policymakers and stakeholders. By adopting data-driven approaches on how countries can most efficiently use resources, decrease emissions, and secure the sustainability of their agriculture in the long run. These efforts are not only crucial for meeting the world's climate goals but also for ensuring food security for future generations.

VI. References (APA Format)

- Cambridge University Press. (2024). Twenty-five rice research priorities for sustainable rice systems by 2050. *Global Sustainability*. https://www.cambridge.org/core/journals/global-sustainability/article/twentyfive-rice-research-priorities-for-sustainable-rice-systems-by-2050/8953AD7E37BE24E365AFEFF704A74641
- Hayek, M. N., Harwatt, H., Ripple, W. J., & Mueller, N. D. (2020). The carbon opportunity cost of animal-sourced food production on land. *Nature Sustainability*, 3(4), 319–325. https://doi.org/10.1038/s41893-020-0483-z
- Le Monde. (2024). Widespread soil degradation alarms UNESCO. *Le Monde*. https://www.lemonde.fr/en/environment/article/2024/07/16/widespread-soil-degradation-alarms-unesco_6685619_114.html
- Reuters. (2024a). From field to the cloud: How AI is helping regenerative agriculture to grow. *Reuters*. https://www.reuters.com/sustainability/land-use-biodiversity/field-cloud-how-ai-is-helping-regenerative-agriculture-grow-2024-09-18/
- Reuters. (2024b). Fertile ground: Why reducing emissions from fertilizer production and use is ripe for innovation. *Reuters*. https://www.reuters.com/sustainability/land-use-biodiversity/fertile-ground-why-reducing-emissions-fertiliser-production-use-is-ripe-2024-10-23/
- Smith, P., et al. (2014). Agriculture, Forestry and Other Land Use (AFOLU). In *Climate Change 2014: Mitigation of Climate Change*. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.
- Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, 108(50), 20260–20264. https://doi.org/10.1073/pnas.1116437108
- Yasmeen, R., Zhao, X., & Jiang, Y. (2022). Assessing agricultural carbon emission efficiency and its driving factors in developing countries. *Environmental Science and Pollution Research International*, 29(22), 33333–33351. https://doi.org/10.1007/s11356-022-19431-4
- Zhang, X., Zhao, H., Wang, J., & Li, Y. (2023). Exploring land-use efficiency and its implications for sustainable agricultural development. *Sustainability*, 15(4), 894. https://doi.org/10.3390/su15041918
- Lobello, A. (n.d.). *Agrofood CO_2 Emission Dataset*. Kaggle. Retrieved from https://www.kaggle.com/datasets/alessandrolobello/agri-food-co2-emission-dataset-forecasting-ml/data
- World Bank. (n.d.). *Agricultural Land Area Dataset*. Retrieved from https://data.worldbank.org/indicator/AG.LND.AGRI.K2?end=2022&start=1990