Data Analysis in Agricultural Sector Enhancement through AI

Presented by Name:

- Eric Maniraguha
  emanirag@andrew.cmu.edu

December 26, 2023

Agenda

- ❑ Problem Statement
- ❑ Key Reason of the project
- ❑ Data Preprocessing
- ❑ Methodology
- ❑ Data Visualization
- ❑ Key Challenges
- ❑ Future Work
- ❑ Conclusion

Problem Statement

In Rwanda, cereal crop productivity is strongly influenced by environmental and seasonal changes. This study zeros in on Season B (March to May) and aims to conduct a data-driven analysis of rainfall patterns and seasonal agricultural data. Focused on cereal crops, our goal is to understand how rainfall impacts yields, spot trends in Seasons B of 2021, 2022, and 2023, and extract insights to enhance farming practices.

01

The findings will refine agriculture strategies and guide informed policy decisions during this crucial rainy season

02

Key reasons for initiating this project

Enhancing Agricultural Productivity

Empowering Farmers

Informed Decision-Making

Policy Formulation

Climate Resilience

Data Preprocessing

Data cleaning

- Checking duplicate records
- Address missing values
- Change Date format
- Data Filtering

Data transformation

- Data Scale up
- Column Transformation

Methodology

01

Data Collection

Utilize NISR agriculture and rainfall dataset.

02

Data Preprocessing

Ensure data quality and standardization for meaningful analysis

03

Exploration and Analysis

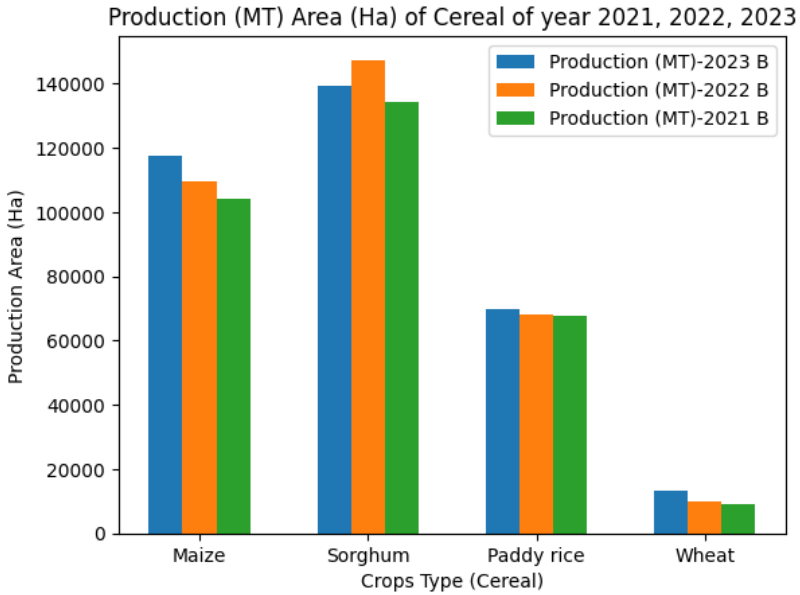
Visualize data patterns using bar diagram.

04

Iterative Process

The methodology is iterative, enabling adjustments based on ongoing findings and stakeholder feedback

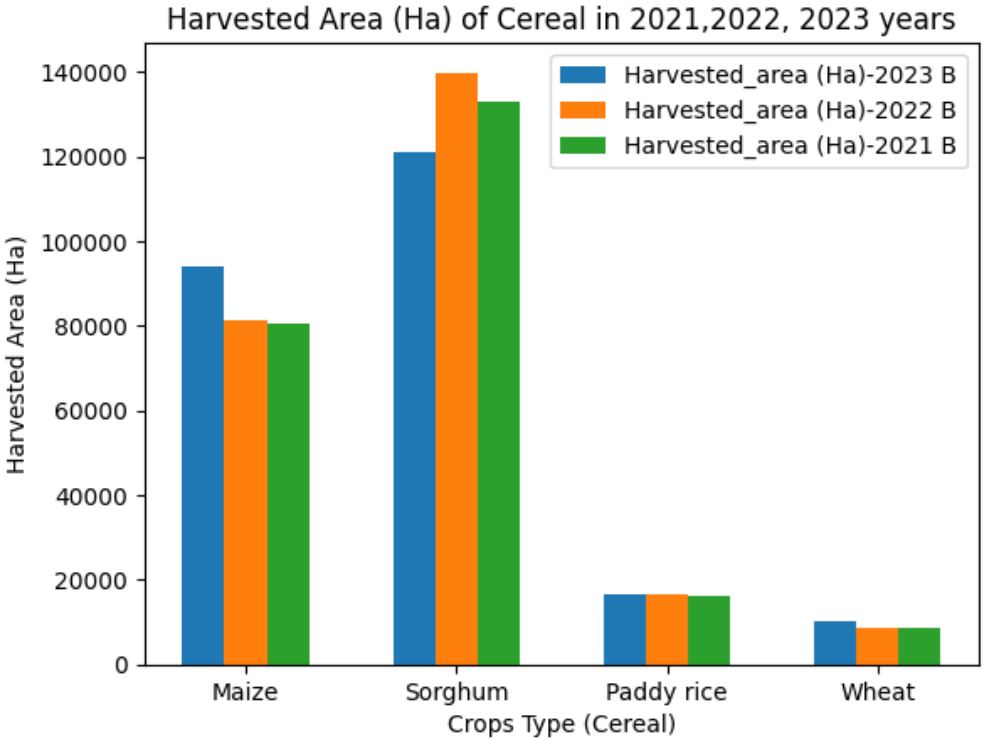
Data Visualization – Production Area



This chart illustrates the production areas of four key cereal crops—maize, sorghum, paddy rice, and wheat—across three consecutive years (2021, 2022, and 2023).

- **Maize:** The highest production area is observed in 2021, reaching 120,000 hectares. This is followed by 2022 and 2023, both falling within the range of 100,000 to 120,000 hectares.
- **Sorghum:** Exhibits a fluctuating trend. The peak production area is in 2022, surpassing 140,000 hectares, followed by a decline in 2023 and 2022, both below 140,000 hectares.
- **Paddy Rice:** Shows consistent production areas over the three years, ranging between 60,000 and 80,000 hectares.
- **Wheat:** Cultivated at a lower production area compared to the other cereals, consistently staying below 20,000 hectares.

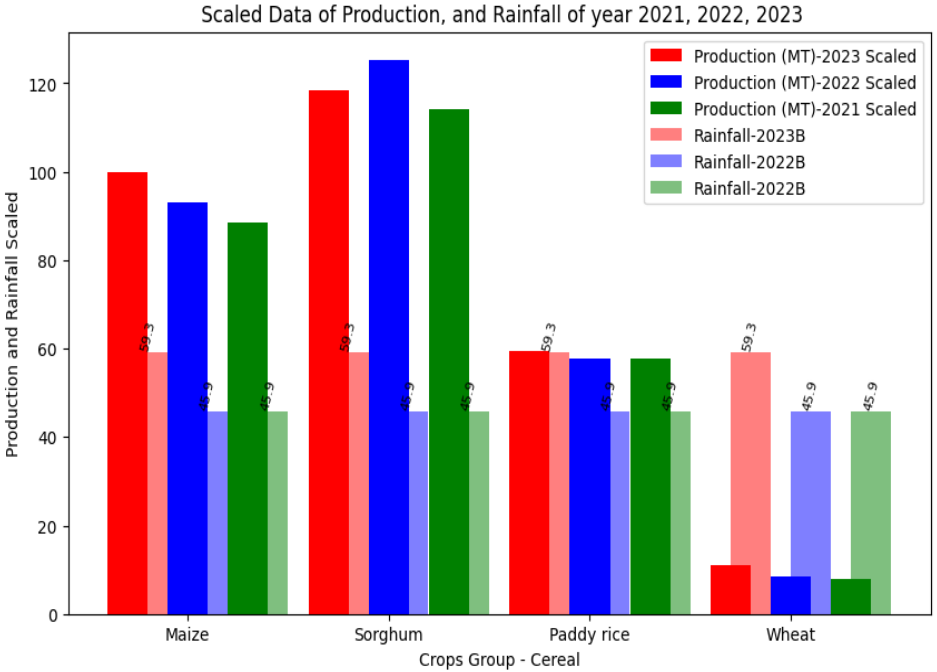
Data Visualization – Haversted Area



This chart visualizes the harvested area of four cereal crops (maize, sorghum, paddy rice, and wheat) across three years (2021, 2022, and 2023).

- **Maize:** Shows the highest harvested area among all crops in 2023, exceeding both 2021 and 2022 levels.
- **Sorghum:** Exhibits a fluctuating trend, with 2022 having the highest harvested area followed by a decline in 2023.
- **Paddy Rice and Wheat:** Consistently maintain lower harvested areas compared to maize and sorghum, both falling below 20,000 hectares throughout the three years.

Data Visualization – Haversted Area



Cereal crops demonstrate unique production and rainfall trends over the three-year period.

- **Maize Dominance in 2023:** Maize takes the lead in production among all crops in 2023, surpassing even paddy rice.
- **Sorghum's Fluctuating Yield:** Sorghum experiences a decline in production in 2023 compared to the preceding year.
- **Steady Rise in Wheat Production:** Wheat exhibits a consistent upward trend in production from 2021 to 2023.
- **Rainfall-Production Discrepancies:** While rainfall levels in 2021B and 2022B remain comparable, crop production varies across these years.
- **2023's Rainfall Dynamics:** Higher rainfall in 2023B aligns with increased maize production but contrasts with reduced sorghum yields.

Key challenges

Data Quality and Availability

Limited access to reliable data, and sensitive farm data might be restricted due to privacy concerns or competition.

Data Preprocessing Complexity

Challenges in cleaning, and standardizing

Multi-Dimensional Data:

Agricultural data is often multi-dimensional, involving spatial and temporal components.

Future work

01 Data Quality Improvement

02 Domain-Specific Advancements

03 Scalability and Efficiency

04 Collaborate with domain experts

Conclusion

- Maize takes the lead in production, showing a substantial rise in 2023.
- Sorghum exhibits fluctuations in production levels.
- Paddy rice and wheat maintain consistent but lower production.
- The study highlights a potential correlation between rainfall patterns and crop production variations.
- Factors such as market demand, climate suitability, and crop resilience may contribute to the observed trends.
- The findings offer crucial insights for farmers, policymakers, and stakeholders to enhance agricultural resilience and productivity in Rwanda.

References

1. **Github Project**: https://github.com/ericmaniraguha/agriculture_data_analysis_rw.git
2. **Rwanda Season Agriculture Survey 2022**: https://microdata.statistics.gov.rw/index.php/catalog/103/study-description
3. **Rwanda Rainfall Dataset:** https://www.fao.org/giews/earthobservation/asis/da

Thank You!