University of Hertfordshire UH

Name : Kumari Duvvari
Student Id : 22097941

Abstract

This study employs an exponential growth model to examine the increase in agricultural land over time. After thorough data processing and feature selection, K-Means clustering reveals distinct patterns among nations. The percentage of Agricultural land, Exports of goods and services, the adjusted net national income growth, the share of agriculture and fisheries in the GDP, and Permanent cropland were selected as important features requiring further investigation. The silhouette score of 0.2169 signifies a robust grouping structure. The primary focus is on understanding the relationship between adjusted net national income growth and the percentage of agricultural land. This research provides valuable insights into the categorization of countries based on economic and agricultural indicators, contributing to a deeper understanding of these dynamics.

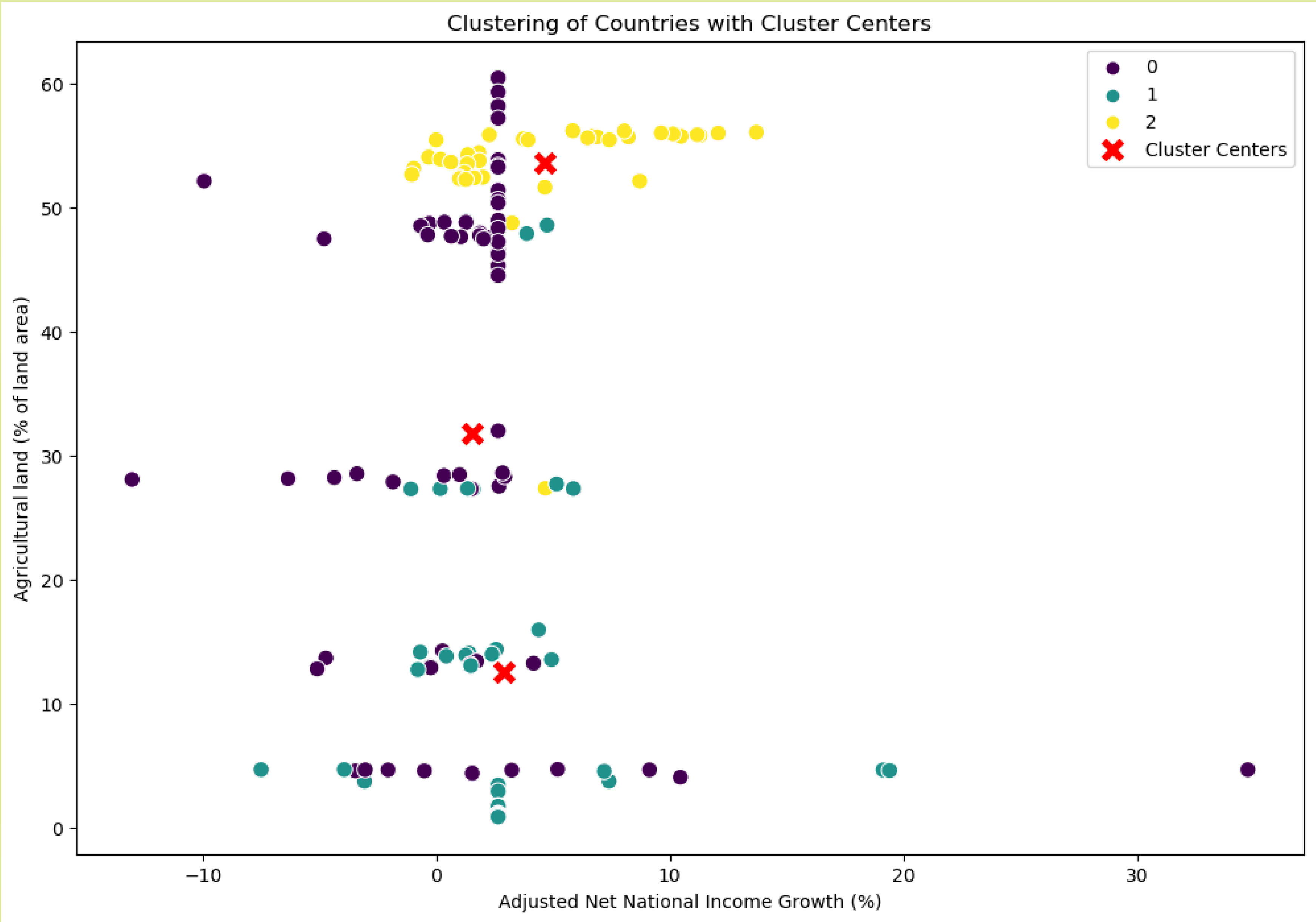
Introduction

Agricultural land is a vital resource that affects food production, the viability of economies, and the health of the environment. This research uses an exponential growth model to simulate the evolution of agricultural land

K-Means Clustering

- **Initialization:**
  - Randomly select K initial cluster centroids.
- **Assignment:**
  - Each data point should be assigned to the cluster with the closest centroid.
- **Update:**
  - Based on the average of the data points in every cluster, recalculate the centroids.
- **Repeat:**
  - Until convergence (when the centroids stabilise or a predetermined number of iterations is reached), repeat the assignment and update stages.

This analysis aims to investigate the relationship between the rise in adjusted net national income and the percentage of land that is agricultural. Furthermore, using certain features, The K-clustering reveals distinct patterns among nations, contributing to a deeper understanding of economic and agricultural dynamics.



Data Pre-processing

Data Loading and Cleaning:
- The Pandas Data Frame was filled with the dataset.
- Values denoted by '..' that were not numeric were substituted with NaN.

Missing Data Handling:
- To ensure that the dataset was full for analysis, simple imputation was used to fill in the missing values for specific columns using the mean value.

Feature Selection:
- A number of pertinent columns were chosen for additional examination, including the proportion of agricultural land, the exports of goods and services, the adjusted net national income growth, the GDP contribution from agriculture and fishery, and the Permanent cropland.

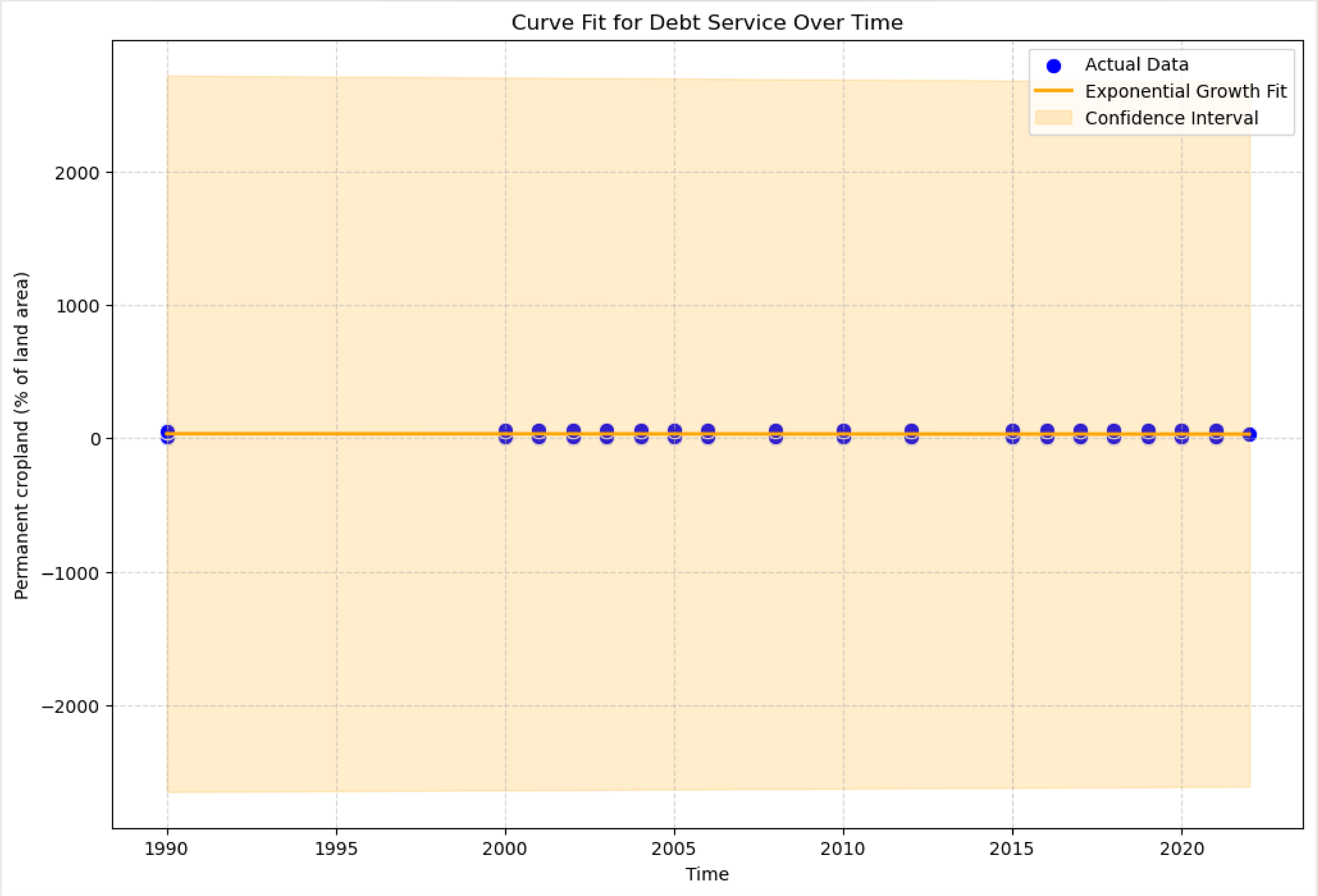
Curve Fitting

The exponential growth model is applied to capture the underlying trend, and predictions are made for future years.

Here is a definition of the selected exponential growth model:

$$\mathcal{Y} = a \cdot e^{bx}$$

Where:
- $\mathcal{Y}$ *is the percentage of arable land.*
- a is a scaling factor.
- b is the growth rate.
- x is the time variable.



The predicted values are as follows:
Predicted value for 2024: 30.74
Predicted value for 2027: 30.47
Predicted value for 2030: 30.21
Predictions for future years (2024, 2027, 2030) are generated using the fitted exponential growth model. This provides a glimpse into the anticipated trends in agricultural land percentage.

Conclusion

The analysis carried out for this paper offers insightful information about the dynamics of agricultural land increase. this analysis serves as a comprehensive exploration of the dynamics between economic growth and agricultural trends. The combination of clustering and modeling techniques provides a robust framework for gaining insights into the complex interplay between economic and agricultural factors on a global scale.

Data source link : https://data.worldbank.org/indicator/AG.LND.AGRI.ZS?view=chart
GitHub link: https://github.com/kumari55/ADS1-Assignment-3.git