University of Hertfordshire UH

Name: Prasad Kiranampalli
Student Id:22073170

Abstract

The modelling of the increase of arable land over time is explored in this work using an exponential growth model. After laboriously loading, cleaning, and imputationing the dataset, a comprehensive dataset was prepared for analysis. The percentage of arable land, the annual freshwater withdrawals, the adjusted net national income growth, the share of agriculture and fisheries in the GDP, and the amount of forest area were selected as important features requiring further investigation. A silhouette score of 0.4650 denotes a strong grouping structure. Using these parameters, the application of K-Means clustering revealed distinct patterns among the nations. Understanding the relationship between the increase in adjusted net national income and the percentage of arable land was the primary objective.

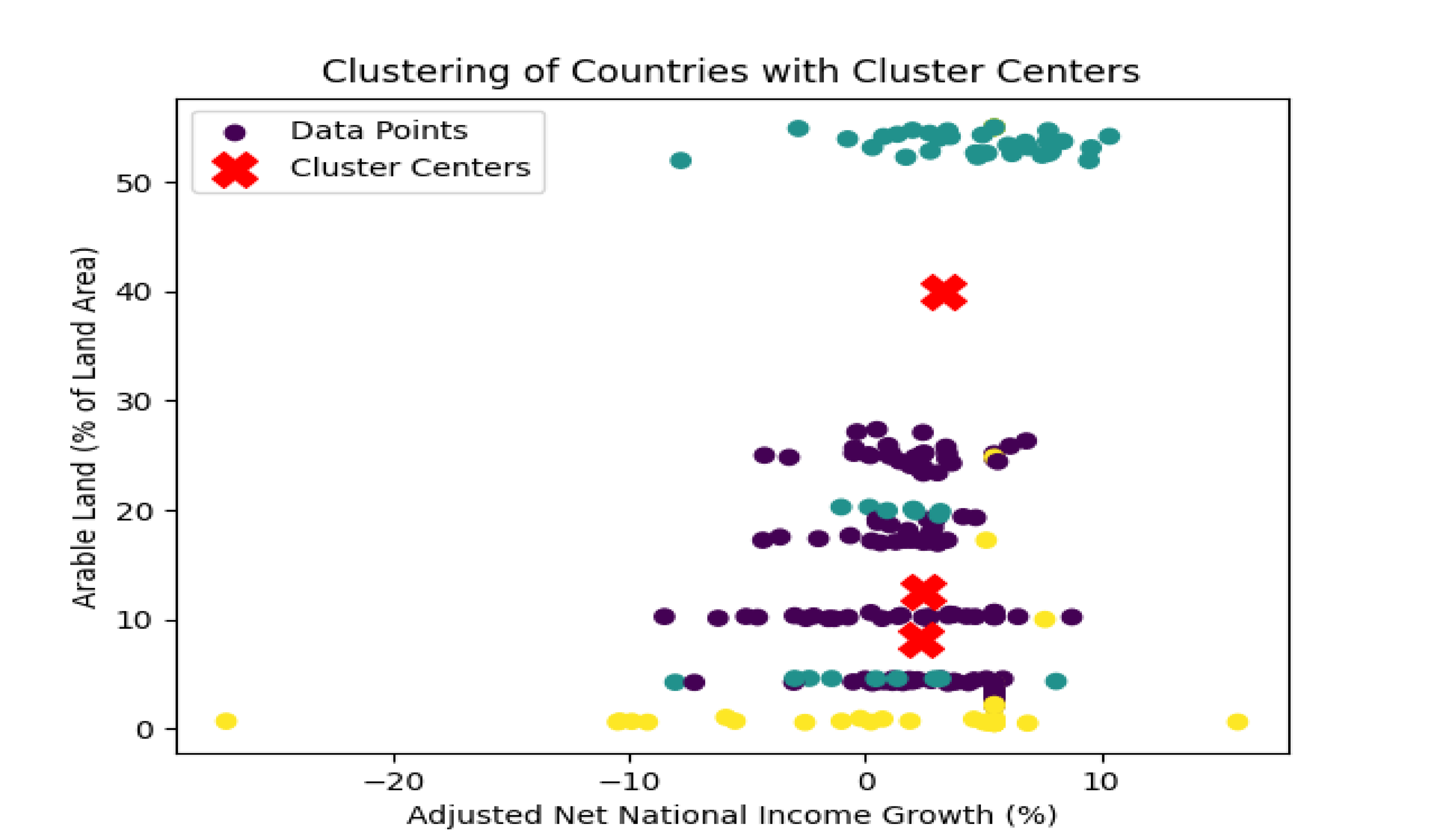
Introduction

Arable 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 arable 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 arable. Furthermore, using certain features, K-means clustering is used to find trends across nations.



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 arable land, the yearly freshwater withdrawals, the adjusted net national income growth, the GDP contribution from agriculture and fishery, and the forest area.

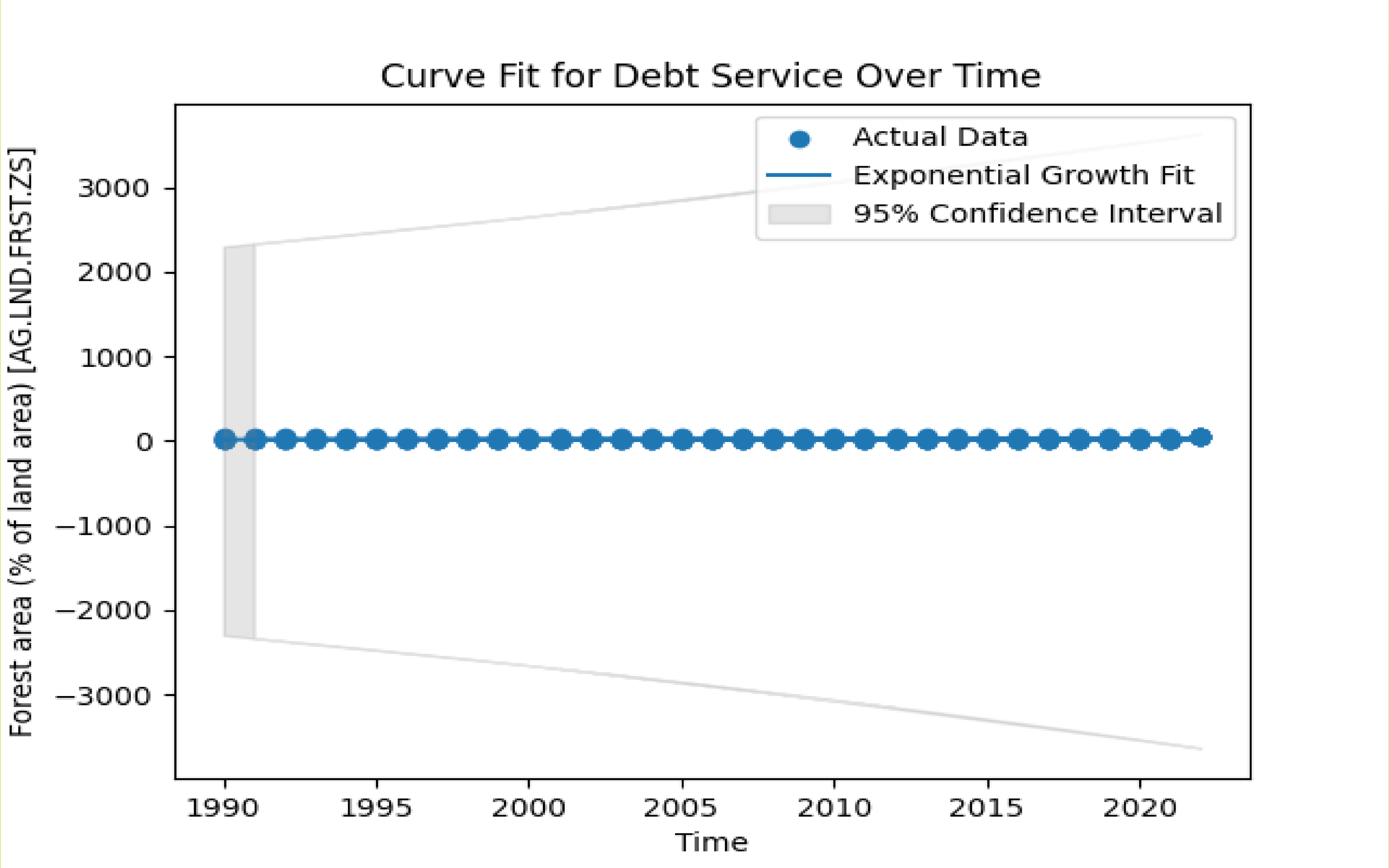
Curve Fitting

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: 21.65
Predicted value for 2025: 21.90
Predicted value for 2026: 22.16

Conclusion

With an emphasis on its importance for food production, economic sustainability, and environmental well-being, the analysis carried out for this paper offers insightful information about the dynamics of arable land increase.

Data source link : https://databank.worldbank.org/reports.aspx?source=2&series=AG.LND.FRST.ZS&country=
GitHub link : https://github.com/kp23aav/ADS1-Assignment-3.git