Analyzing The Distribution Of Arable Land And Forest Area Across The World Using Clustering And Fitting Techniques

University of Hertfordshire

School of Physics, Engineering and Computer Science

By Hathamunage Harsha Sandaruwan (22031386)

Abstract

The aim of this study was to analyze the distribution of arable land and forest area across various countries using clustering and fitting techniques. The dataset used in this study was obtained from the World Bank and included indicators such as arable land and forest area. The data was preprocessed and then clustered using k-means clustering to identify groups of countries with similar arable land and forest area characteristics.

Subsequently, we performed fitting using arable land data to model the relationship between arable land and forest area. The results showed that the countries with the highest arable land and forest area were clustered together, and the countries with the lowest values were also clustered together.

Furthermore, the fitted model suggested that there was a negative relationship between arable land and forest area, meaning that as arable land increased, forest area decreased. This study provides insights into the distribution of arable land and forest area across countries and can inform policy decisions related to sustainable land use and conservation efforts.

Introduction

The World Bank provides a wealth of data on various indicators related to different countries around the world. In this study, we focused on the arable land and forest area indicators. Arable land is defined as land under temporary crops (double-cropped areas are counted once), temporary meadows for mowing or pasture, land under market or kitchen gardens, and land temporarily fallow. Forest area, on the other hand, refers to land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems (for example, in fruit plantations and agroforestry systems) and trees in urban parks and gardens.

The objective of this study was to cluster countries based on their arable land and forest area data using a clustering algorithm. The clustering results were then used to fit a regression model to predict the arable land indicator based on the forest area indicator. This analysis can help in understanding the relationship between the two indicators and provide insights into how they are related in different countries. The findings of this study can have implications for policymakers and stakeholders who are interested in sustainable land use practices and conservation efforts.

Objectives

- To analyze the distribution of arable land and forest land across countries using clustering techniques.
- To identify patterns and trends in the data that could be useful for policymakers, such as countries with high levels of deforestation or countries with underutilized arable land.
- To develop a model that fits the arable land data and could be used to predict future trends in arable land use.
- To assess the relationship between arable land and forest land and identify any potential conflicts or trade-offs between these two land uses.

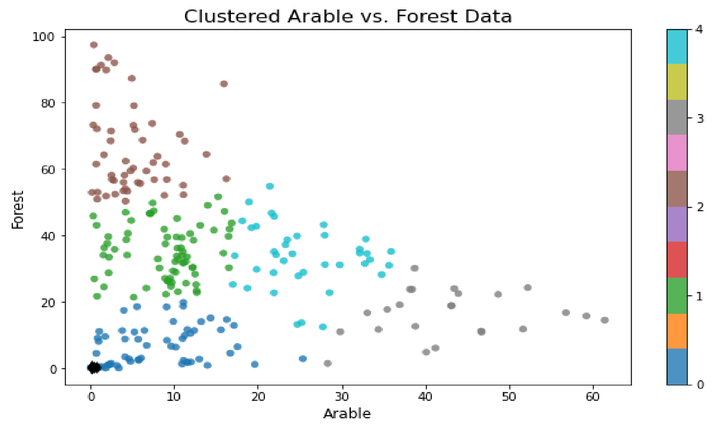
Dataset

The dataset used in this analysis is sourced from the World Bank and contains information on the indicators of arable land and forest area for various countries. Arable land refers to land that is suitable for agricultural cultivation, while forest area refers to land that is covered by trees with a minimum height of 5 meters and a canopy cover of at least 10%. The dataset is containing information on these indicators for multiple years, and for different countries across the world. The dataset is useful for understanding the trends in land use and the potential impacts of land use changes on the environment.

University of Hertfordshire UH

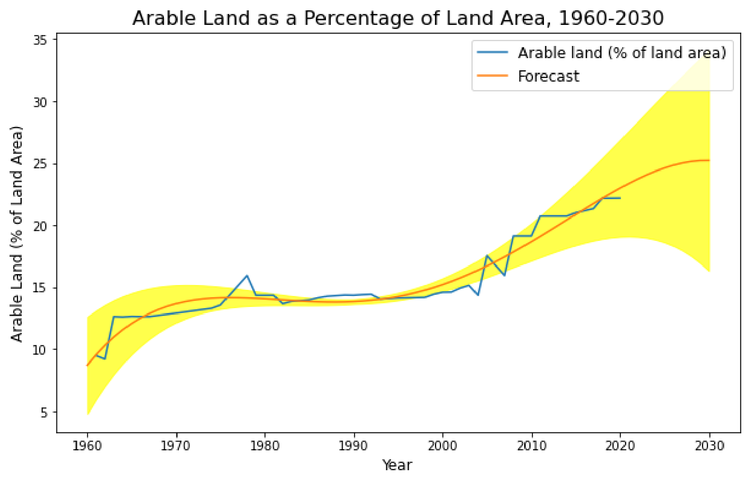
Findings Based On Analyse

As per the below clustered graph we can find that countries in grey cluster have a high percentage of arable land and a low percentage of forest area, while countries in brown cluster have a low percentage of arable land and a high percentage of forest area. This information could be useful for understanding how different countries are using their land resources and for identifying potential areas for conservation or agricultural development.



Making A Prediction For Sri Lanka Using Fitting Technique

Based on below prediction graph, it seems that the percentage of arable land in Sri Lanka has been increasing slowly over the years. Therefore, we can make a prediction as Sri Lanka's arable land percentage will be increase up to 25% of the total land by 2030. However, the rate of increase has been inconsistent and sometimes even decreasing. Factors such as government policies, and economic development can all impact the rate of increase or decrease in arable land. It would be best to consider these factors in future analysis to make an accurate prediction.



Conclusion

When looking at the clustering of countries based on their arable land and forested area, it is possible to identify 5 main groups. The first group includes countries with a high percentage of both arable land and forested area, such as Denmark and Slovenia. The fifth group includes countries with a low percentage of both arable land and forested area, such as Kuwait and Qatar. Overall, it is clear that the distribution of arable land and forested area across countries is complex and varies widely. Understanding these patterns is important for policymakers, as it can help inform decisions about land use, conservation, and agricultural development.

References

GitHub Repository Link - https://github.com/Harsha-hathamunage/ADS1---Assignment-3-Clustering-and-fitting/tree/main

GitHub Repository Link to 2nd Dataset - https://github.com/Harsha-hathamunage/ADS1---Assignment-3-Clustering-and-fitting/blob/main/SL_Arable.csv

Hathamunage Harsha Sandaruwan
hh22abm@herts.ac.uk