

ABSTRACT:

This investigation uses K-means clustering on a variety of datasets to investigate clustering trends in environmental and agricultural variables. A few examples of the indicators that have been chosen include "Agricultural land (% of land area)," "Access to clean fuels and technologies for cooking," and "Forest area (% of land area)." The findings show clear patterns among the nations, offering perceptions on prospective trends and difficulties in the region. Furthermore, curve fitting is applied in a concentrated case study on India to simulate the exponential expansion of the forest area, providing insightful information for environmental sustainability and policy choices.

Introduction:

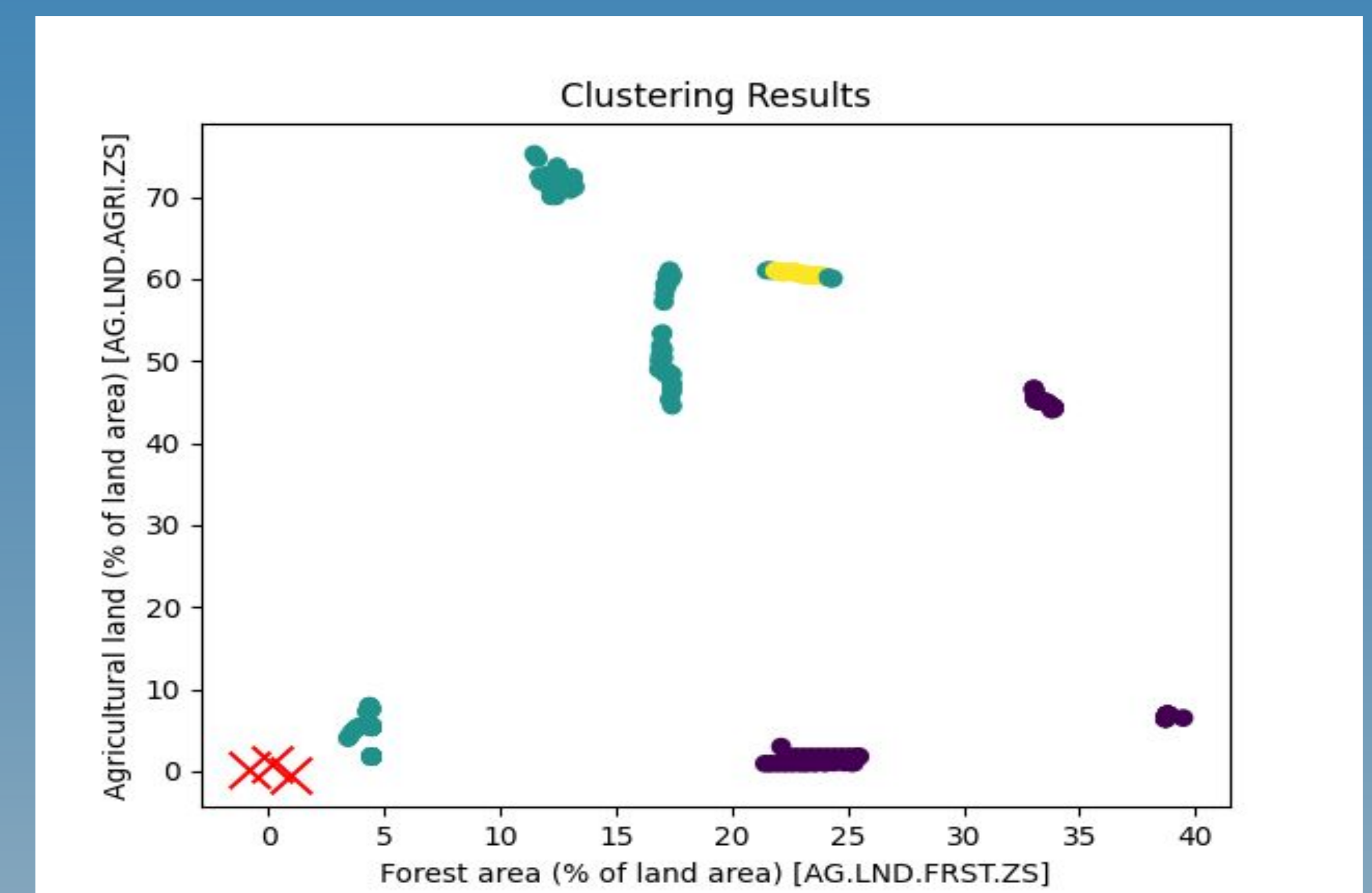
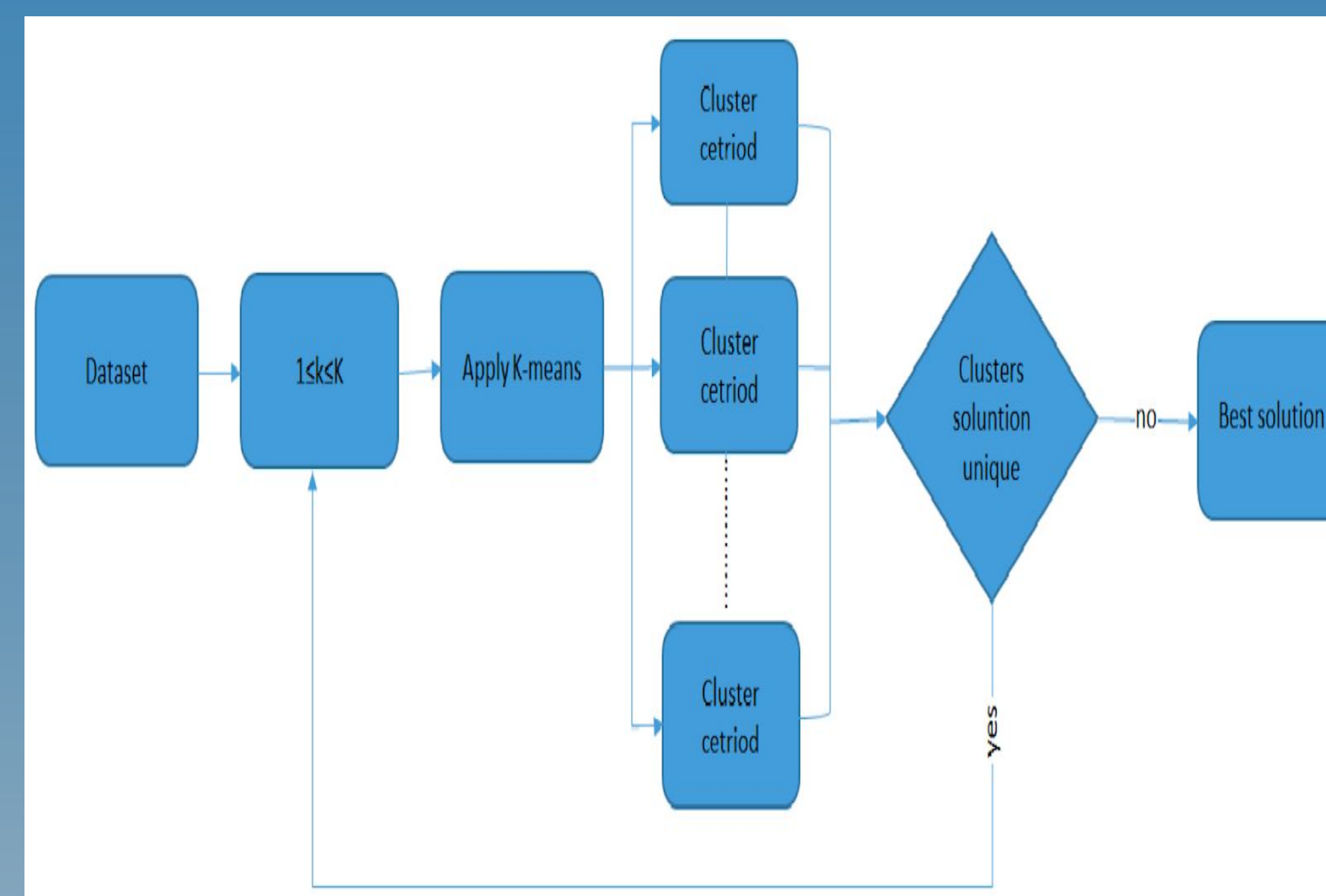
This analysis's goal is to investigate the patterns of clustering in a dataset pertaining to agricultural and environmental variables. Furthermore, our goal is to fit an exponential growth model to a particular time series in the dataset in order to reveal possible future trends.

Data Pre-processing:

To enable accurate analysis, the dataset was loaded and preprocessed. Using the mode of each column, missing values were imputed in pertinent columns, such as "Forest area (% of land area)".

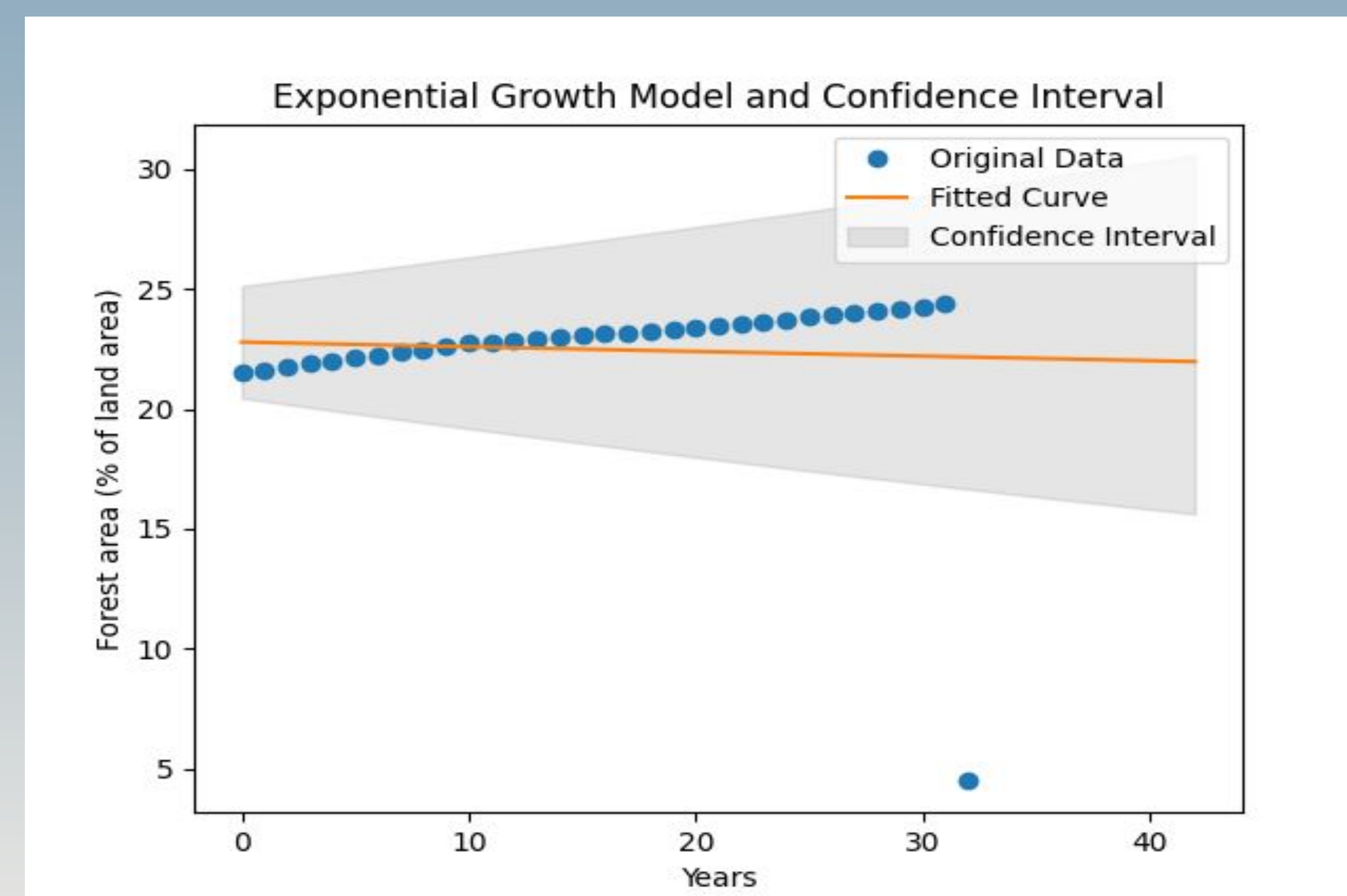
Clustering:

Using the normalised values of the chosen indicators, the K-means clustering algorithm was used to find patterns and classify the countries. In this investigation, three clusters were selected. With the x-axis denoting "Forest area (% of land area)" and the y-axis denoting "Agricultural land (% of land area)," the scatter plot below shows the clustering results. Red "x" icons designate the cores of each cluster, offering information about its central trends. Achieved a silhouette score of 0.4809.



Curve Fitting:

In order to investigate the temporal trends in "Forest area (% of land area)," a curve fitting technique was utilised. The time series data for India were fitted with an exponential growth model. The graphic below illustrates the curve fitting results, which include the fitted curve, the original data points, and confidence intervals:



For Australia the plot is as follows:

