

Clustering and Fitting

Abstract

This study used K-Means clustering to examine data pertaining to agriculture and forests in order to identify unique trends across nations. The three clusters that were established provided significant insights for both economic and environmental concerns, since they represented regions with differing degrees of agricultural and forest features. A forward-looking viewpoint was offered by time series analysis and forecasts for the trends in forest areas in China, Pakistan, and Australia. By enhancing our understanding of the complex interactions between forestry and agriculture, the study helps stakeholders and policymakers make decisions that will support sustainable development.

Introduction

Data pertaining to agriculture and forestry are the main focus of the analysis, which also looks for trends, groups patterns, and forecasts for certain nations. The dataset utilised includes statistics on arable land, agricultural land, and forest area for a number of nations and years.

Data Loading and Cleaning

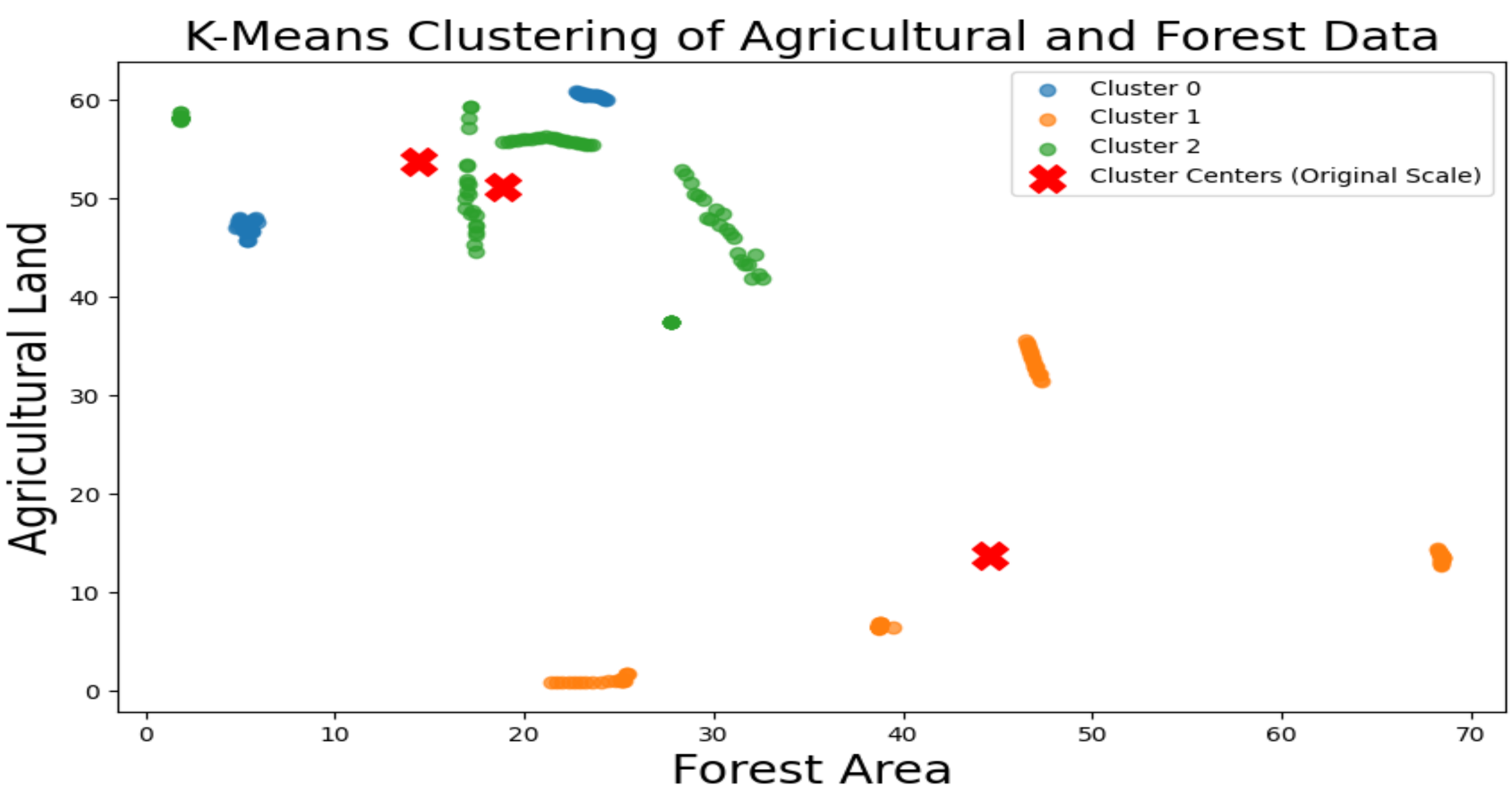
After pre-processing processes were completed, the data was loaded from a CSV file. Data was transposed for simplicity of analysis, and missing values in numerical columns were imputed using the mean. A few pertinent columns were chosen for clustering, such as "Arable Land," "Agricultural Land," and "Forest Area."

Clustering Analysis

The dataset's patterns were found using K-Means clustering. Based on insights or requirements from the company, three clusters were selected. Prior to performing clustering, the data was normalised using Standard Scaler. The quality of the clustering was evaluated by computing the silhouette score, which is a measure of cluster cohesiveness and separation. The properties of each feature's mean value within the clusters were analysed and presented.

Insights

- Cluster 0:** Distinguished by a high share of arable land and a comparatively low amount of wooded area. This cluster can stand for areas with a high concentration of agricultural activity and little forest cover.
- Cluster 1:** Has a large percentage of forest area and a small amount of agricultural land. This cluster indicates areas that may have less agricultural development and a substantial amount of forest cover.
- Cluster 2:** Balances land use between agriculture and forest cover, signifying areas with a significant proportion of agricultural activity and a moderate amount of forest cover.

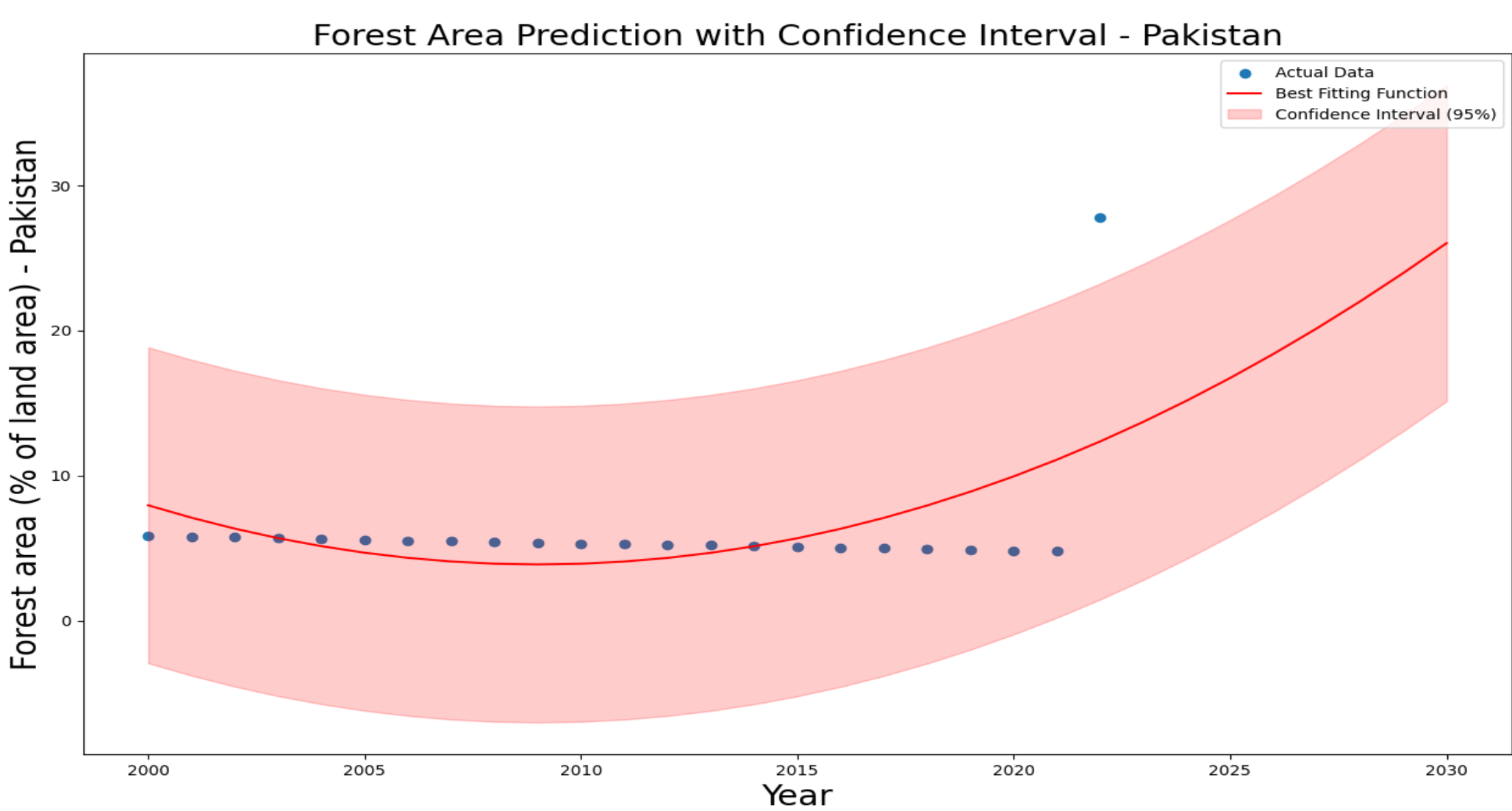


Time Series Analysis and Prediction

A time series analysis of the forest area data was carried out for a subset of the countries, including Australia, China, and Pakistan. The historical data was fitted using a quadratic function, and curve fitting and error range calculations were used to calculate confidence intervals (95%). This sheds light on the historical pattern of the forest area and the uncertainty that goes along with it.

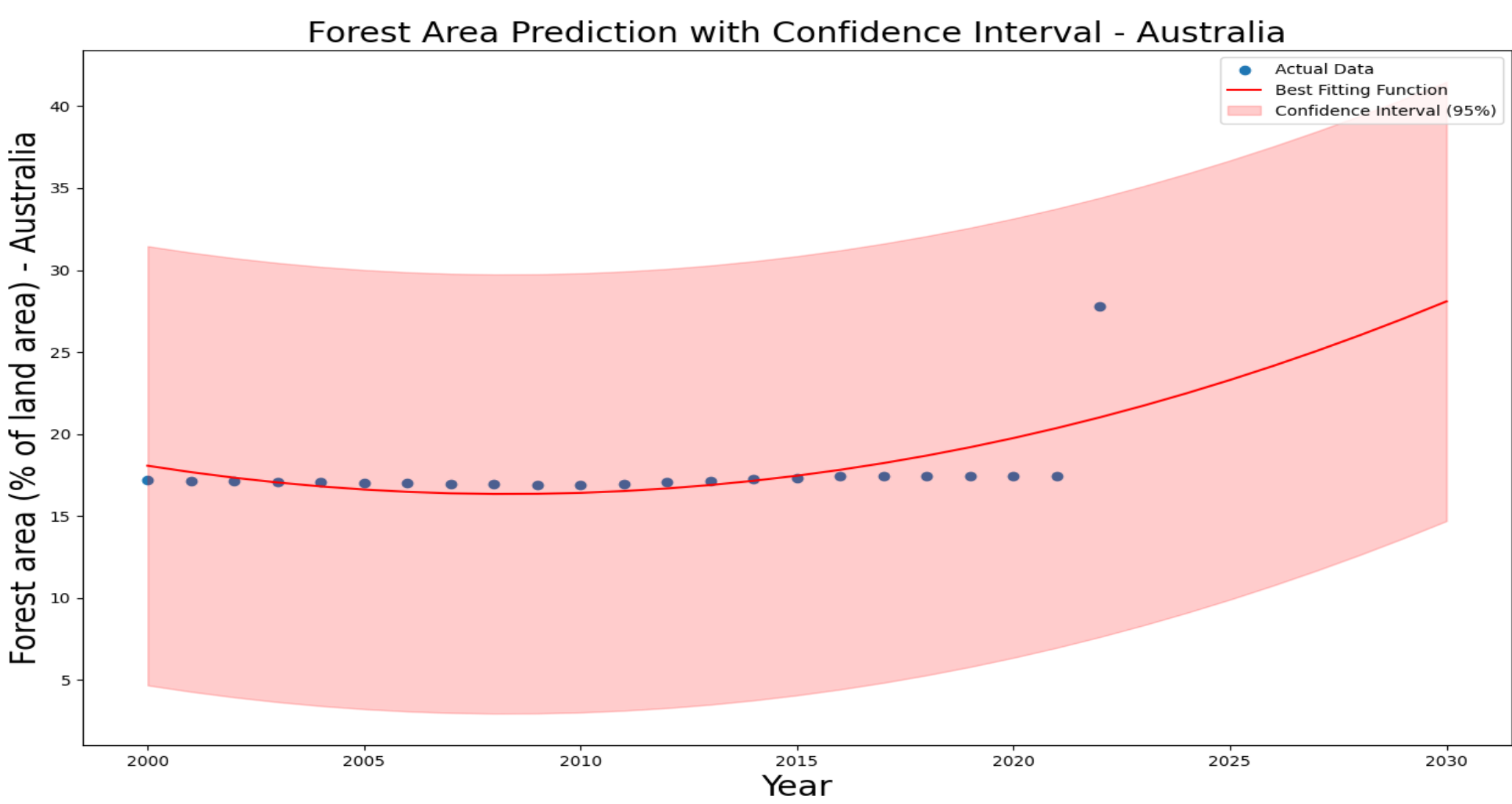
Predictions for Pakistan:

Pakistan's forest area will grow over the years, from 2020 to 2030.



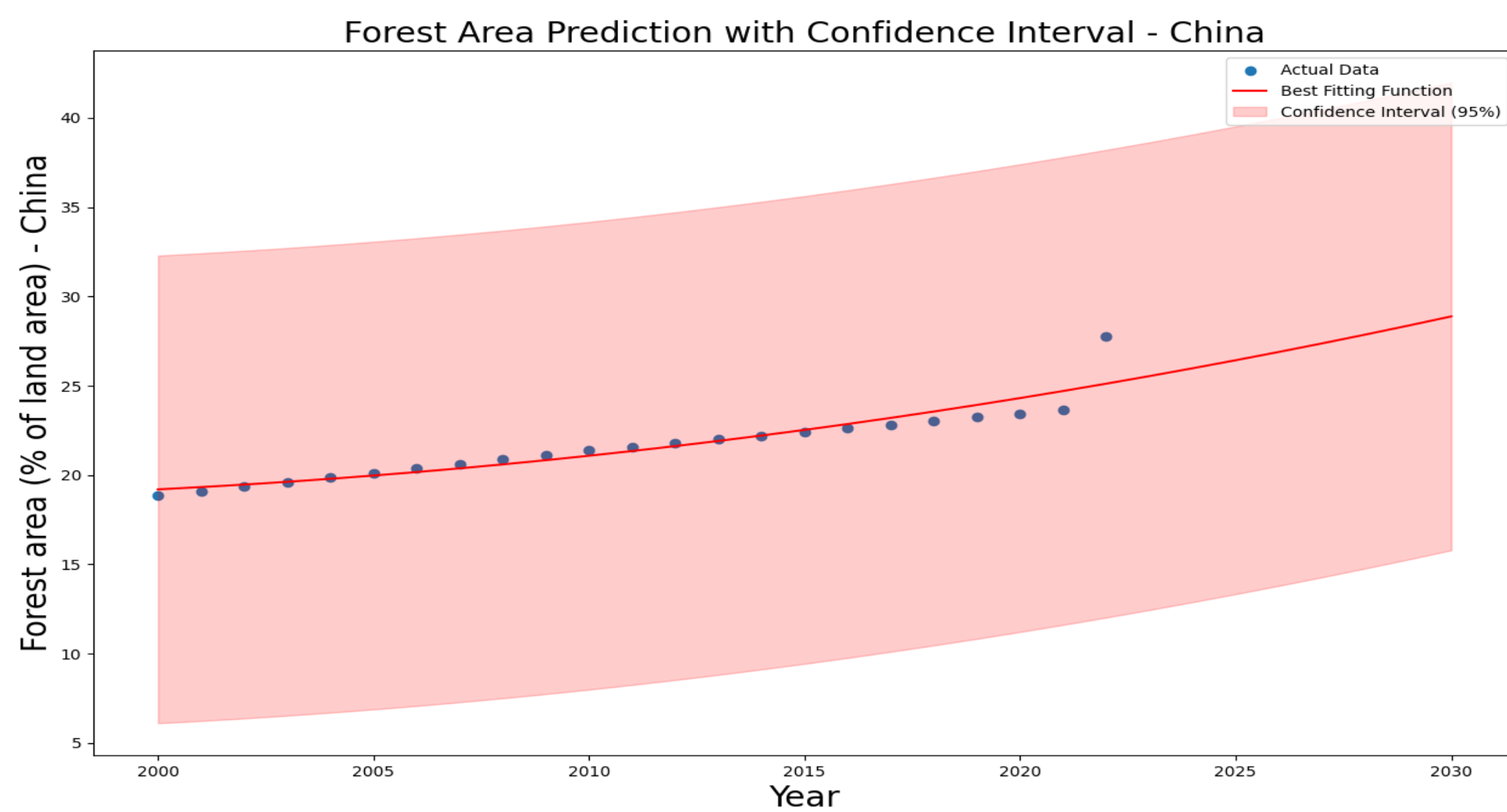
Predictions for Australia:

Australia's forest acreage is growing yearly as well.



Predictions for China:

The amount of forest land will only slightly grow between 2020 and 2030.



Conclusion:

We used K-Means clustering to find unique patterns among different nations in this extensive study of data pertaining to agriculture and forests. By dividing the land area into three clusters—forest, agricultural, and arable—it became clearer which areas had which ecological and agricultural traits. Cluster 0 showed places with a high concentration of agricultural activity and a low covering of forests, indicating areas that are largely dependent on agriculture. Cluster 1 demonstrated areas with high levels of forest cover and little agricultural development, emphasizing the significance of conservation initiatives. Cluster 2 portrayed a landscape in equilibrium, with notable agricultural activity and a moderate amount of forest cover. In addition, a time series analysis concentrating on trends in forest area was carried out for a few chosen nations, including China, Australia, and Pakistan. We supplied forecasts and confidence ranges for the next using quadratic curve fitting.

Student Id: 22069870
Name : Ajay Prabhat Gorrumuchu

Datasource:
<https://databank.worldbank.org/reports.aspx?source=2&series=AG.LN.D.FRST.ZS&country=>

Githubsource:
<https://github.com/ajayprabhatGorrumuchu/ClusteringAndFitting.git>