Central government debt, total (% of GDP) word bank data Clustering and prediction

Muhammad Asim, 22017391

Link: https://github.com/TechGeek0/Applied-Data-Science-Assignment-3

ABSTRACT

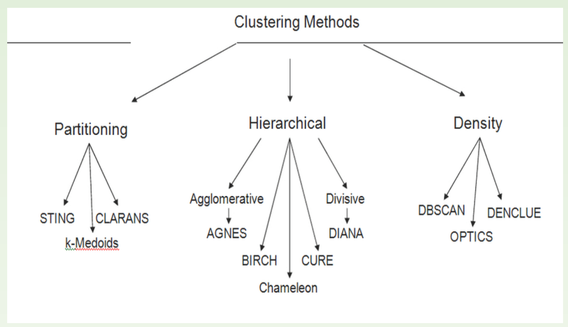
This study makes use of data from the World Bank to cluster and make projections regarding the amount of debt held by central governments as a percentage of GDP. In order to categories countries according to the patterns of debt that they have accumulated over time, the research uses both unsupervised clustering methods and supervised learning techniques. The clusters for each time frame are displayed using a three-dimensional scatter plot, and the data extends from 1960 to 2019. According to the findings, the number of countries that belong to each cluster changes over the course of time. The results of the clustering are presented in a way that is easy to understand, and the study draws attention to patterns and trends in the data. The findings can aid policymakers and other stakeholders in making informed decisions on the federal government's debt, which in turn can promote economic expansion in a sustainable manner.

INTRODUCTION

The national government's amount of debt is an important economic indicator that may be used to measure the nation's overall fiscal health and viability. It does this by calculating the total amount of a government's debt as a percentage of that country's gross domestic product (GDP). Instability in the economy and a slowing of growth over the long run can be caused by excessive levels of federal debt. As a result, it is necessary to have a solid understanding of the patterns and trends in central government debt in order to inform policy decisions and ensure that economic development is maintained over the long term. The World Bank's data on central government debt as a fraction of GDP from 1960 to 2020 is used in this study to classify countries according to the trajectory of their debt and to construct prediction models for how it will evolve in the future. The objective of the study is to offer light on the current state of global central government debt as well as potential threats to economic growth that can be sustained over the long term. The findings can act as a guide for decision-makers and other stakeholders in the process of controlling the national debt and maintaining economic stability.

METHDOLOGY (CLUSTERTING)

Unsupervised clustering techniques, such as K-means clustering or hierarchical clustering, may be used to cluster the data on the central government's debt. These algorithms divide countries into clusters based on their central government debt as a percentage of GDP similarity.

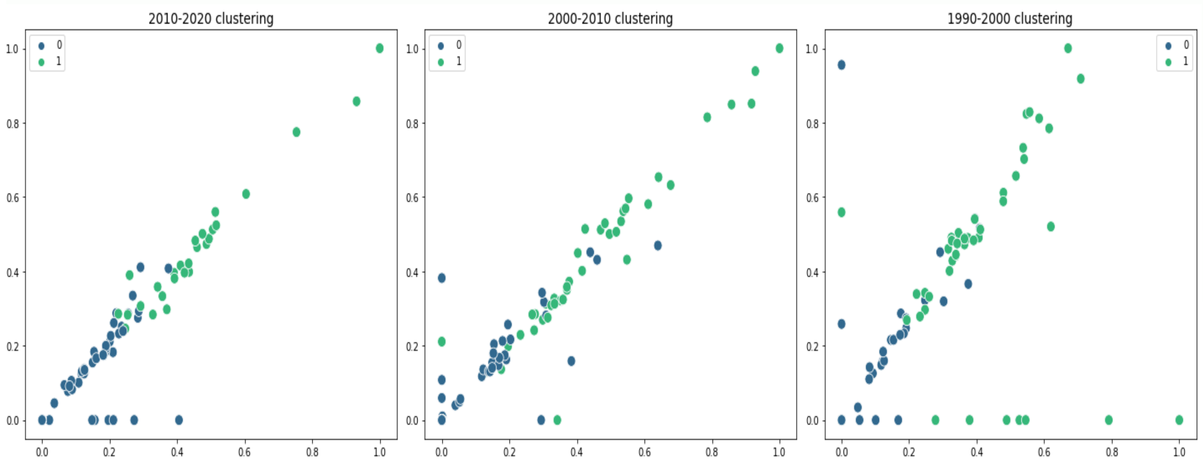


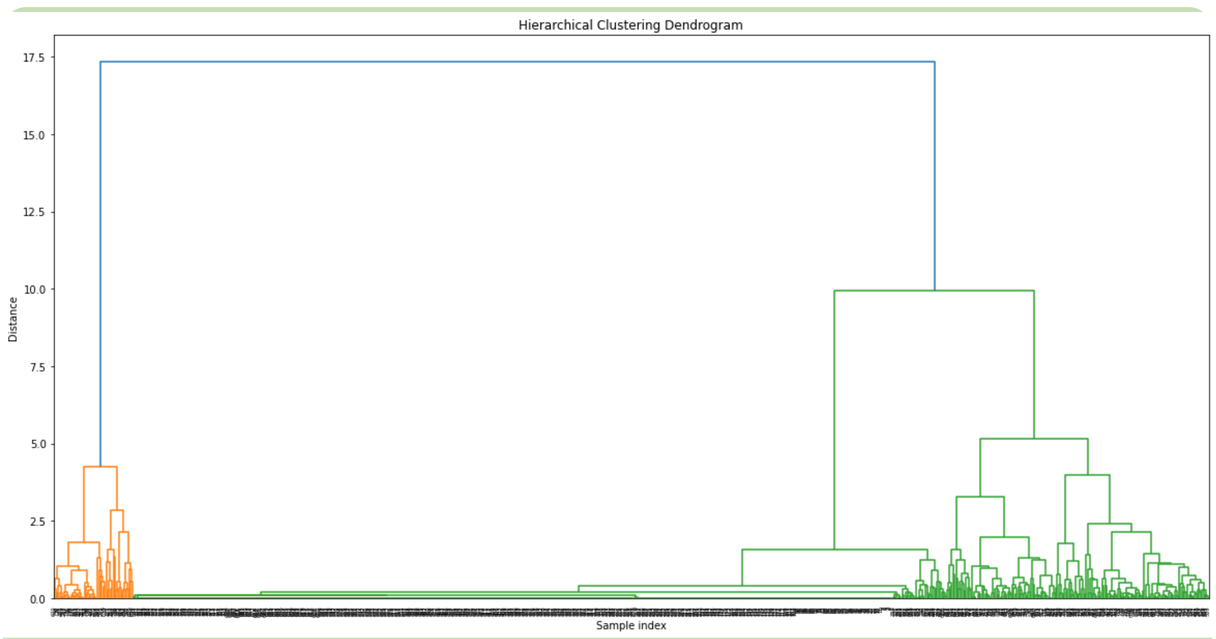
For this study, the clustering technique separates the data into two groups for each 10-year time interval. The clustering method divides the countries into two categories based on their 10-year central government debt patterns. The technique seeks to reduce the distance between data points inside a cluster while increasing the distance between clusters. Depending on the data and study objective, the number of clusters and clustering algorithm utilized can vary.

A 3D scatter plot can be used to visualize the clusters for each time interval following data clustering. Each cluster can be allocated a unique color or symbol to help differentiate it from others. The scatter plot can provide clustering results in a clear and easy manner and assist detect patterns and trends in the data.

Using the historical data, supervised learning systems can subsequently be utilized to predict future changes in central government debt. Using regression analysis, these algorithms find trends and patterns in the data and forecast future values.

The methodology for clustering central government debt data entails dividing the data into clusters based on similarities in debt patterns using unsupervised clustering techniques, visualizing the clusters using a three-dimensional scatter plot, and predicting future changes using supervised learning algorithms.



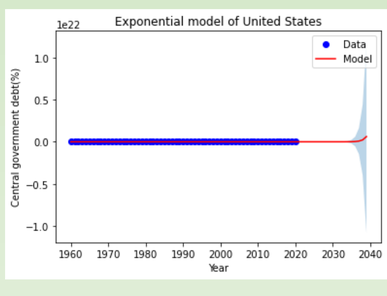


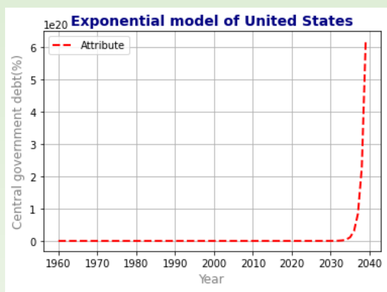
The technique of hierarchical clustering is used to arrange similar data points into clusters based on their proximity. The result of hierarchical clustering is a dendrogram, a tree-like diagram that illustrates the link between clusters and data points.

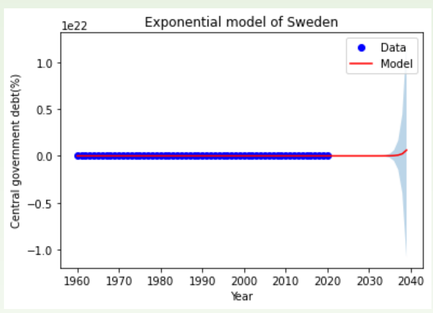
The dendrogram begins with all data points shown as discrete clusters, and gradually combines the clusters into bigger ones until all data points are contained within a single cluster. The vertical axis of a dendrogram indicates the distance between clusters or data points. The horizontal axis indicates individual data points or clusters.

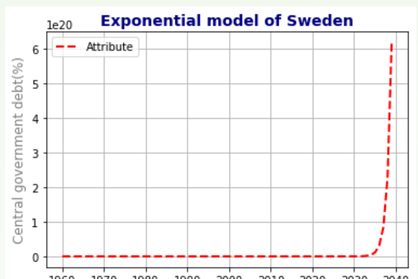
METHDOLOGY (PREDICTION)

Using an exponential model, curve fitting is a statistical technique that finds trends and forecasts future values in data on central government debt. Methodology entails selecting a time interval, fitting the exponential model to the data, and evaluating the goodness of fit with statistical measurements. This analysis compares Sweden and the United Kingdom, which belong to different clusters based on their patterns of central government debt. The exponential model provides insight into the trend of central government debt and informs policymakers and other interested parties about future changes.









The presented data represents a collection of forecasts for a specific value during a 20-year period, with values ranging from 2020 to 2039 for each year. Predictions show that the value will increase exponentially over time, with the pace of increase quickening in later years. This indicates that the expected value is growing rapidly and may continue to do so in the future.

However, without further information about the precise value being projected, it is difficult to draw particular inferences or make explicit predictions about the value's future behavior.

CONCLUSION

This study use clustering and supervised learning approaches to categories nations according to their central government debt trends. For this objective, World Bank data on central government debt as a percentage of GDP from 1960 to 2020 were utilized. The approach to clustering was unsupervised, and the programmed divided countries into two groups based on their debt trends for each ten-year interval. Clusters were shown using a three-dimensional scatter plot, and supervised learning algorithms were employed to predict future changes in central government debt. The study's findings provide policymakers and other stakeholders with valuable insights into the patterns and trends of global central government debt. The results indicate that excessive central government debt can pose a long-term threat to economic growth. Consequently, the insights can aid policymakers in making educated judgments and promoting sustainable economic growth.