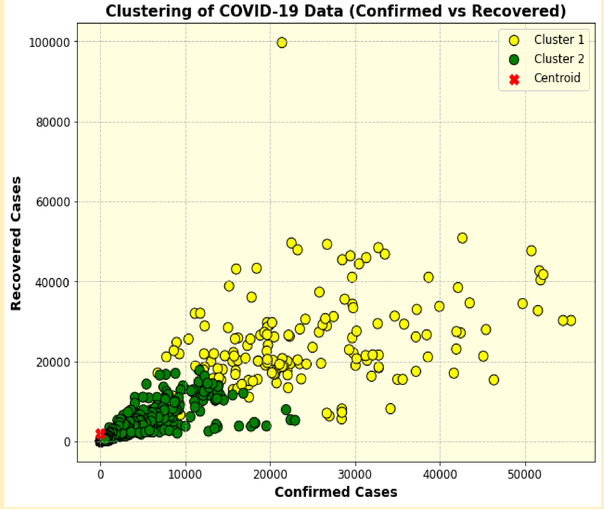
Clustering and Fitting Analysis of COVID-19

Abstract

This poster presents a concise analysis of COVID-19 data, focusing on K-means clustering and a modified exponential growth model. Key findings include distinct patterns in confirmed and death cases through clustering, revealing temporal trends over the years. The modified exponential growth model provides valuable insights into forecasting confirmed cases.

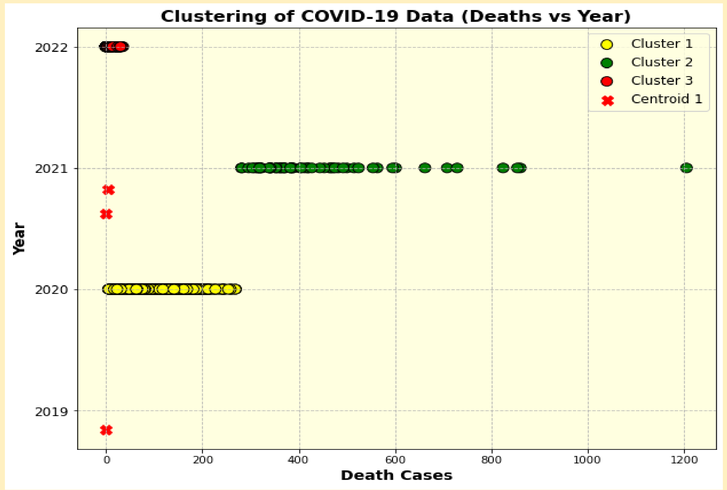
Introduction

Amid the COVID-19 pandemic, this poster aims to swiftly communicate pivotal findings from a clustering and modelling analysis of COVID-19 data. Through K-means clustering, two main clusters in confirmed vs. recovered and death cases are identified. Temporal analysis reveals three clusters in death cases across 2020-2022..



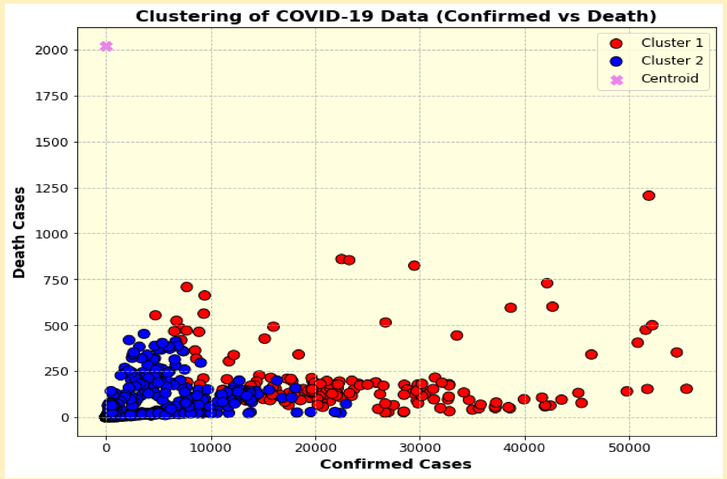
Clustering of Confirmed vs Recovered Cases

This plot shows the confirmed and recovered. cases. Two main clusters identified, Centroids marked for each cluster. Data points are grouped into two distinct clusters, highlighting different patterns in the relationship between confirmed and recovered cases, we use scatter plot to for the graph. Use different colours for each cluster to enhance visual distinction.



Clustering of Death Cases vs Year

Data points are categorized into three distinct clusters, indicating variations in the relationship between death cases and the corresponding years (2020, 2021, 2022). Centroids, representing the average values of death cases within each cluster, are visually highlighted on the plot.



Confirmed vs Death Cases

clustering is a powerful technique that groups similar data points into clusters, providing insights into underlying trends. Centroids, representing the centre points of each cluster, are marked on the plot. These centroids act as central tendencies for the respective clusters, providing a reference point for the average characteristics of the grouped data.

Growth Model Fitting for Confirmed Cases

This graph illustrates the application of a modified exponential growth model to the observed confirmed COVID-19 cases. The model is fitted to the data, providing insights into the trend, and a confidence interval is plotted to indicate the uncertainty associated with the model. A curve representing the best-fit of the modified exponential growth model to the observed confirmed cases. The curve captures the underlying trend and growth pattern.

Conclusion:

K-means clustering identifies patterns in COVID-19 data, particularly in the confirmed and recovered cases relationship. Temporal analysis of death cases across 2020, 2021, and 2022 exposes evolving patterns. A modified exponential growth model provides a quantitative framework for forecasting confirmed cases. Policymakers can tailor strategies based on identified clusters, addressing unique characteristics. The integration of clustering, temporal analysis, and modelling provides a holistic understanding of COVID-19 data. The study contributes valuable knowledge to public health by unravelling patterns, trends, and forecasting capabilities.