TECHNICAL REPORT ON COVID 19 DATASET

By FIDELIS ENEOJO ISAH

FELLOW ID: FE/23/8804427 (COHORT)

Introduction

This report analyzes the COVID-19 dataset, including data exploration, visualization, and predictive modeling. The dataset comprises daily records of confirmed cases, deaths, and recoveries across various countries.

Data Exploration

The initial step involved loading and cleaning the dataset to ensure accuracy and consistency. Key statistics such as the total number of cases, deaths, and recoveries were computed. The data was then segmented by country and date to identify trends and patterns.

Data Visualization

To gain insights into the spread and impact of COVID-19, several visualizations were created:

Heatmaps: Showed the geographical distribution of cases, highlighting hotspots and regions with significant outbreaks.

Point plots: illustrated individual data points, trends, patterns, comparisons, and distributions between variables.

Bar Charts: Compared the total number of cases, deaths, and recoveries across different countries.

Predictive Modeling

Predictive models were developed to forecast future trends in COVID-19 cases. The following models were employed:

Linear Regression: Used to predict the number of cases based on historical data.

Time Series Analysis (ARIMA): Applied to model and forecast the time-dependent nature of the data.

Machine Learning Models (e.g., Decision Tree): Enhanced prediction accuracy by capturing complex patterns in the data.

Results

The analysis revealed several key findings:

Trend Analysis: Identified periods of rapid increase and decline in cases, correlating with government interventions and public health measures.

Geographical Insights: Highlighted regions with high transmission rates and the effectiveness of containment strategies.

Model Performance: The predictive models demonstrated varying degrees of accuracy, with machine learning models (SKLearn) providing the most reliable forecasts.

Conclusion

The analysis of the COVID-19 dataset provided valuable insights into the pandemic's progression and impact. The visualizations and predictive models can aid in decision-making and resource allocation for future outbreaks. Continuous monitoring and updating of the models are essential to maintain their accuracy and relevance.