Vivekanand Education Society's
Institute of Technology
(Autonomous Institute Affiliated to University of Mumbai, Approved by AICTE & Recognised by Govt. of Maharashtra)
NAAC accredited with 'A' grade

Semesteí: VI Subject : AIDS - 1

Title of the Project :
Covid-19 Statistical Analysis and Prediction using ARIMA model

Domain: Statistical Forecasting Techniques & Prediction Modelling

Member 1:Dev Gaonkar D15C 12
Member 2: Advik Hegde D15C 15
Member 3: Shreyash Kamat D15C 22

Professor Name : Dr. Ravita Mishra

Content

- Overview
- Introduction
- Problem Statement
- Objectives
- Literature Survey
- Proposed System
- Proposed Design
- Implementation
- Conclusion
- References

Introduction to Project

The COVID-19 pandemic had widespread and unprecedented effects globally, with countries like India being particularly hard-hit. India experienced multiple waves of infections, which put immense pressure on the healthcare system and led to significant mortality rates. To understand the severity and predict future trends, it is crucial to analyze the available data on COVID-19 cases, deaths, and other relevant factors. This project focuses on analyzing these trends and predicting future death counts in India using the ARIMA (AutoRegressive Integrated Moving Average) model, a powerful time-series forecasting tool.

Problem Statement

Despite the availability of COVID-19 data, effective forecasting methods are needed to help policymakers and healthcare professionals prepare for future waves.
Challenges include:

- High variability in cases and deaths across regions
- Lack of real-time prediction for healthcare planning
- Difficulty in selecting the right forecasting model for accurate trend predictions

Using time series modeling (ARIMA), we aim to compare historical and predicted death rates, aiding in proactive decision-making.

Objectives of the project

- To analyze COVID-19 data trends in India.
- To implement ARIMA for time-series forecasting of COVID-19 death counts.
- To visualize results through line graphs showing actual vs. predicted deaths.
- To provide insights for policymakers and the healthcare sector.

Literature Survey

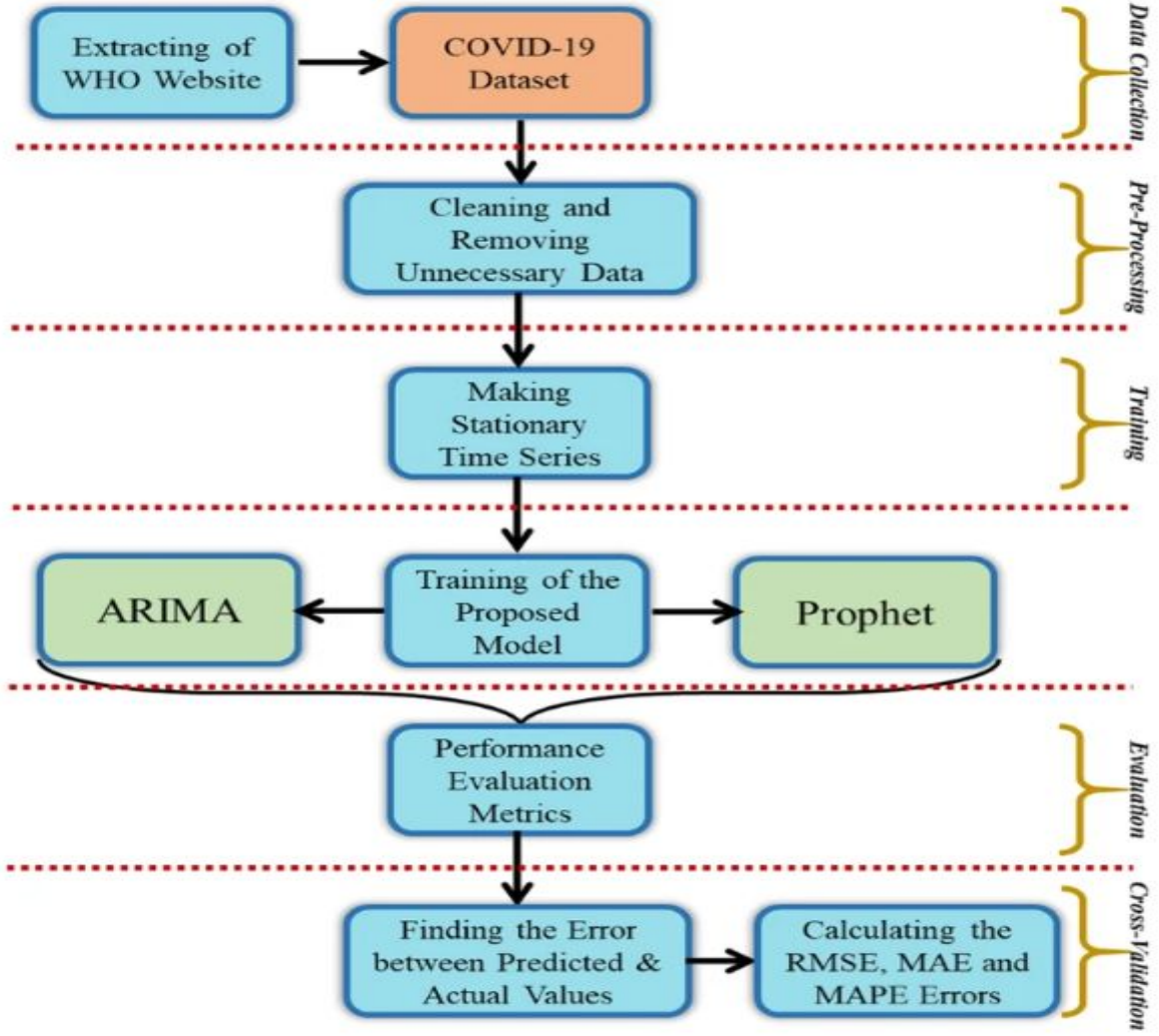
Sr.No	Title	Author	Publish Date/Year	Description
1	An Application of ARIMA Model to Forecast the Dynamics of COVID-19 Epidemic in India	Rupinder Katoch, Arpit Sidhu	2021	This study applies the ARIMA model to forecast the progression of the COVID-19 epidemic in India. By analyzing the trends of confirmed cases and deaths, the authors aim to provide insights into the future trajectory of the pandemic, which can assist policymakers in planning and resource allocation.
2.	Predictive Analysis of COVID-19 Using LSTM and ARIMA Models with Various Orders in India	Subhalaxmi Chakraborty, Arindam Chakraborty, Pritam Kundu,University of Engineering and Management, Kolkata, India	2022	This research compares the effectiveness of LSTM and ARIMA models in predicting COVID-19 cases in India. The authors analyze various orders of the ARIMA model to determine the most accurate forecasting approach, providing valuable insights for understanding the pandemic's potential future impact.

Proposed System

- Data Collection & Cleaning – Preprocessing the Kaggle dataset on COVID-19 in India.
- Feature Selection – Extracting relevant data: date, state-wise cases, recovered, deaths.
- Model Training – Implementing ARIMA for time-series forecasting.
- Model Evaluation – Analyzing MAE, RMSE, and MAPE for accuracy.
- Visualization & Interpretation – Comparing predicted vs. actual death counts.

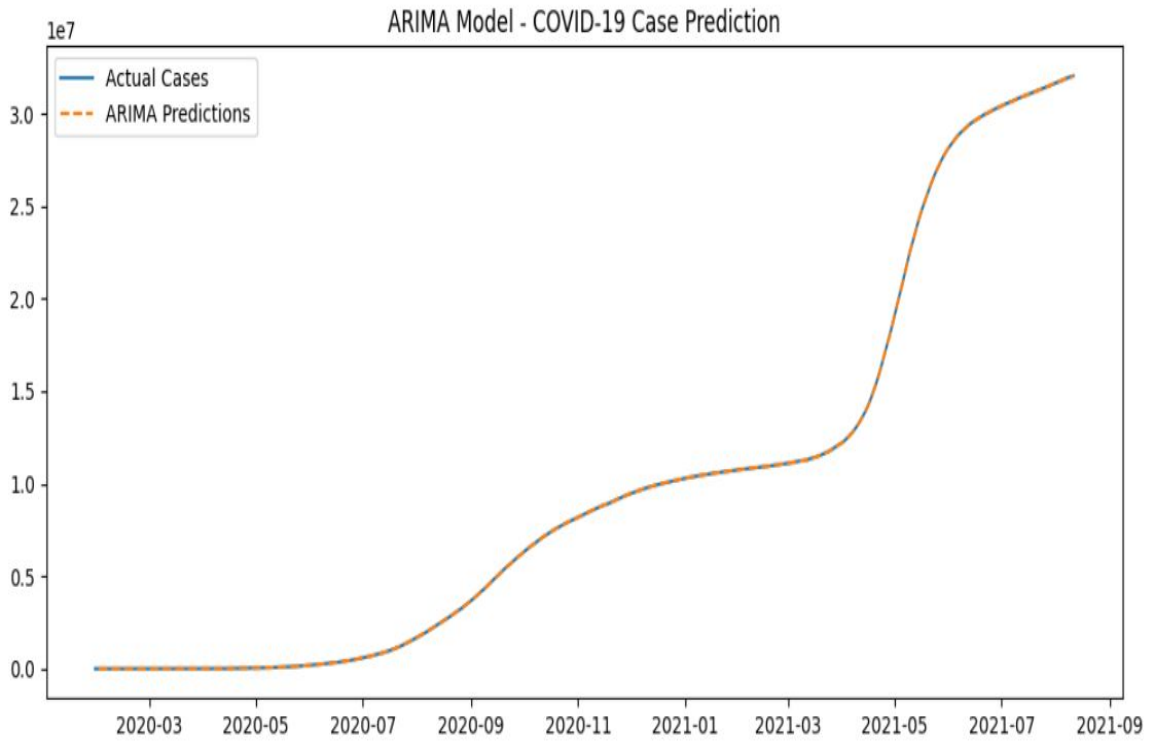
Proposed Design

Fig. 1



Implementation

1. Predicted V/S Actual Values – The forecasted values by the ARIMA model.



The ARIMA model is able to capture the trends within the dataset effectively.It performs much better than a simple linear regression model or even a polynomial regression model.

Linear Regression Model:
Mean Absolute Error: 3230706.39
RMSE: 3846176.55

Polynomial Regression Model:
Mean Absolute Error: 9706844.62
RMSE: 10120034.75

Arima Model Values :
Mean Absolute Error: 4320.16RMSE: 7608.46

Implementation

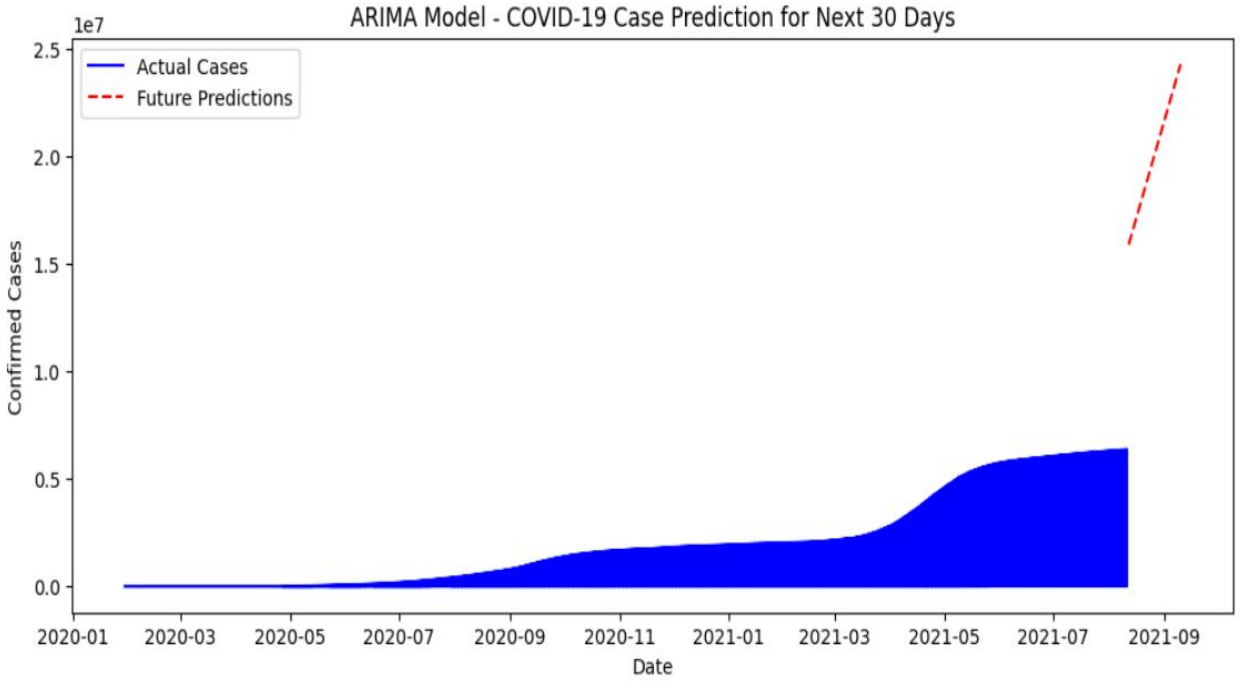
2. A comparison drawn between the states to figure out relative performance

The UT of Dadra and Nagar Haveli emerged as the best performing state based on the ratio of Deaths-to-Recovery.

```
Best-performing state based on Death-to-Recovery Ratio:
State/UnionTerritory       Dadra and Nagar Haveli
Confirmed                                   20722
Cured                                       20352
Deaths                                          8
Recovery Rate (%)                       98.214458
Fatality Rate (%)                        0.038606
Death-to-Recovery Ratio                  0.000393
Name: 9, dtype: object
```

Implementation

3. A graph comparing latest recorded values to predicted values for the next 30 days.



The graph shows the latest recorded values for confirmed cases and shows next to it the predicted or expected trend of values based on the trends captured from the dataset by using the ARIMA model.

Conclusion

This ARIMA model effectively forecasts short-term COVID-19 death trends in India.The predictions assist policymakers in resource planning and healthcare preparedness,especially for preparation for future outbreaks and pandemic risks.The ARIMA model works best for short-term forecasts and assumes a stable trend but loses accuracy if external factors influence results heavily and if the model is not consistently trained on new data.Incorporating deep learning models (LSTM) can hugely improve accuracy and offer more precise results.

The results from this model could be possibly extended and aid in deciding protocols and policies for future outbreaks.

Additional data from other sources like vaccination results could also be incorporated to expand the scope, usability and reliability of this model.

References

- Kaggle Dataset: COVID-19 in India
- https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-022-07472-6
- https://www.nature.com/articles/s41598-022-06218-3

Links to the reference papers

1. Research Paper 1 - Link 1
2. Research Paper 2 - Link 2