Minor Project

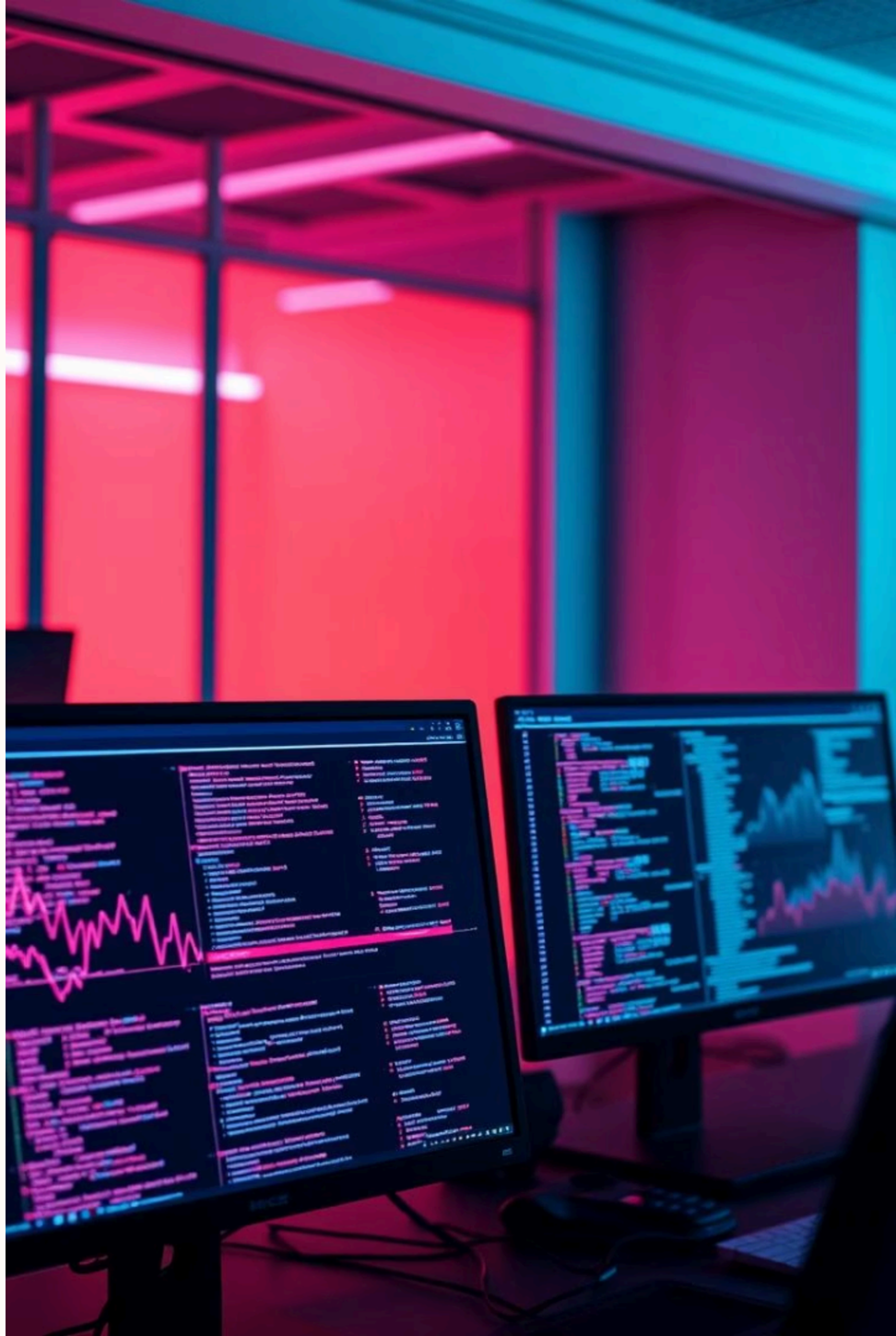
Name – Hariom (0901AM221037)

Kanishk Parihar (0901AM221039)

Subject – minor project

Time Series Analysis and Anomaly Detection BY using ARIMA

Predictive analytics on time series data to enable accurate forecasting and real-time anomaly detection.



The Power of Time Series Analysis

What is Time Series Analysis?

Study of trends, seasonality, and cycles in data over time.

Its Importance

Crucial for forecasting and detecting anomalies in many sectors.

Applications

Finance, healthcare, manufacturing, and many more industries.

Project Overview

Objectives and methods for robust predictive analytics.

Project Objective: Forecasting & Anomaly Detection

Develop Forecasting Model

Build a robust ARIMA model for precise predictions.

Detect Anomalies

Implement real-time statistical anomaly detection techniques.

Create User Interface

Design intuitive visualizations for data insights.

Evaluate Performance

Use RMSE and MAE metrics to assess model accuracy.

Tools & Technologies Used

Programming

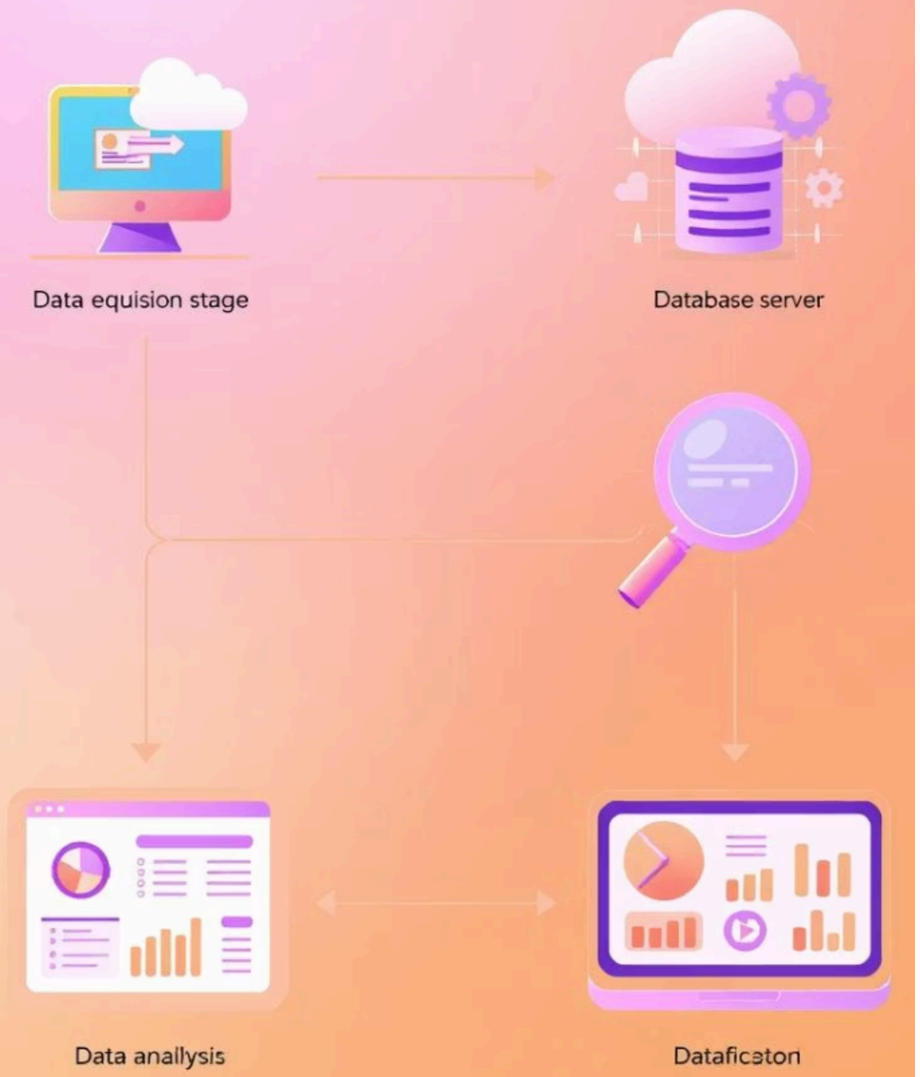
- Python 3.9+
- Pandas, NumPy for data analysis
- Statsmodels for ARIMA modeling

Machine Learning & Visualization

- Scikit-learn algorithms
- Matplotlib, Seaborn for charts
- Flask for web app development

Development Environment

- Jupyter Notebook
- Visual Studio Code



System Architecture: Data Flow and Components

Data Acquisition

Scripts fetch and store time series from APIs or databases.

Preprocessing

Cleaning, transforming, and engineering features for modeling.

Modeling & Detection

ARIMA forecasting combined with statistical anomaly detection.

Visualization & Deployment

Interactive dashboards powered by Flask for real-time insights.

Data Acquisition & Preprocessing

Data Sources

Example: Stock prices via the Yahoo Finance API.

Ensures reliable, continuous data flow.

Cleaning & Transformation

- Handle missing data and outliers
- Create lagged variables and rolling stats
- Apply differencing for stationarity

Data Splitting

Training (70%), validation (15%), and testing (15%) sets.

Time Series Forecasting – ARIMA Model

ARIMA combines autoregression, differencing, and moving averages.

1. Select optimal (p, d, q) via ACF and PACF plots.
1. Train on historical data to capture patterns.
1. Validate model to ensure prediction accuracy.
1. Use equations ARIMA(p, d, q) to model time dependencies.

Anomaly Detection: Identifying Outliers

Statistical Methods

Z-score, modified Z-score, and IQR to detect anomalies.

Thresholding

Set thresholds based on statistical properties.

Real-Time Detection

Algorithms to analyze and flag anomalies instantly.

Visualization

Anomalies shown clearly on interactive time series plots.

Visualization & Output: Interactive Dashboards

User Interface

Developed using Flask for accessible web interaction.

Charts

Display forecasts and anomalies with interactive exploration.

Outputs

Export data as CSVs, generate reports, and trigger alerts.

Results, Discussion & Use Cases

Model Accuracy

Strong RMSE and MAE results on test data.

High precision and recall in anomaly detection.

Use Cases

- Predictive Maintenance in Manufacturing
- Fraud Detection in Finance
- Demand Forecasting in Retail
- Healthcare Monitoring

Conclusion

Project achieved robust forecasting and real-time alerts.

Future work includes expanding datasets and model enhancements.

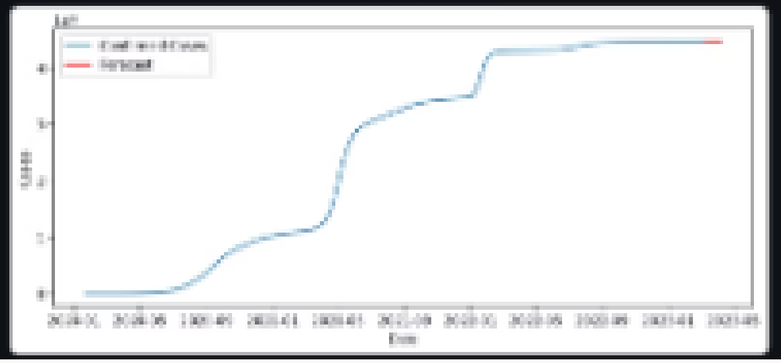
COVID-19 Disease Surveillance Dashboard

India COVID-19 Stats

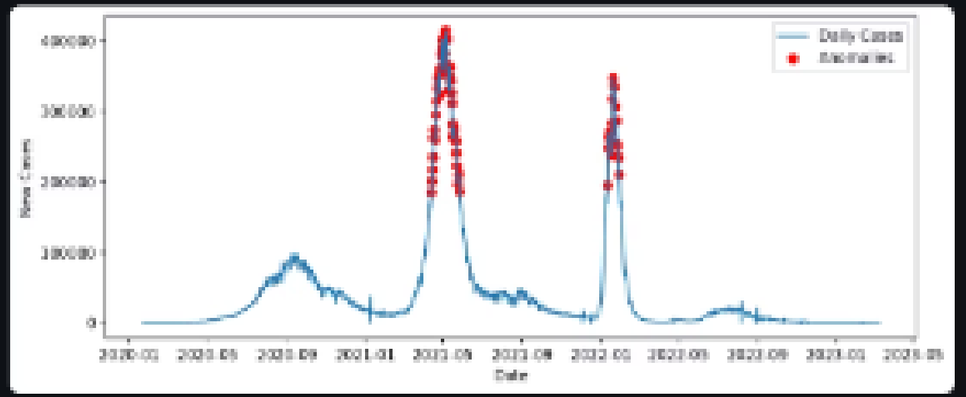
Total Confirmed Cases: 4480728

Total Confirmed Deaths: 138779

Confirmed Cases + Forecast (30 days)



Daily New Cases with Anomalies



Forecast Table

Date	Forecasted Cases
2023-03-10 00:00:00	44591300
2023-03-11 00:00:00	44592700