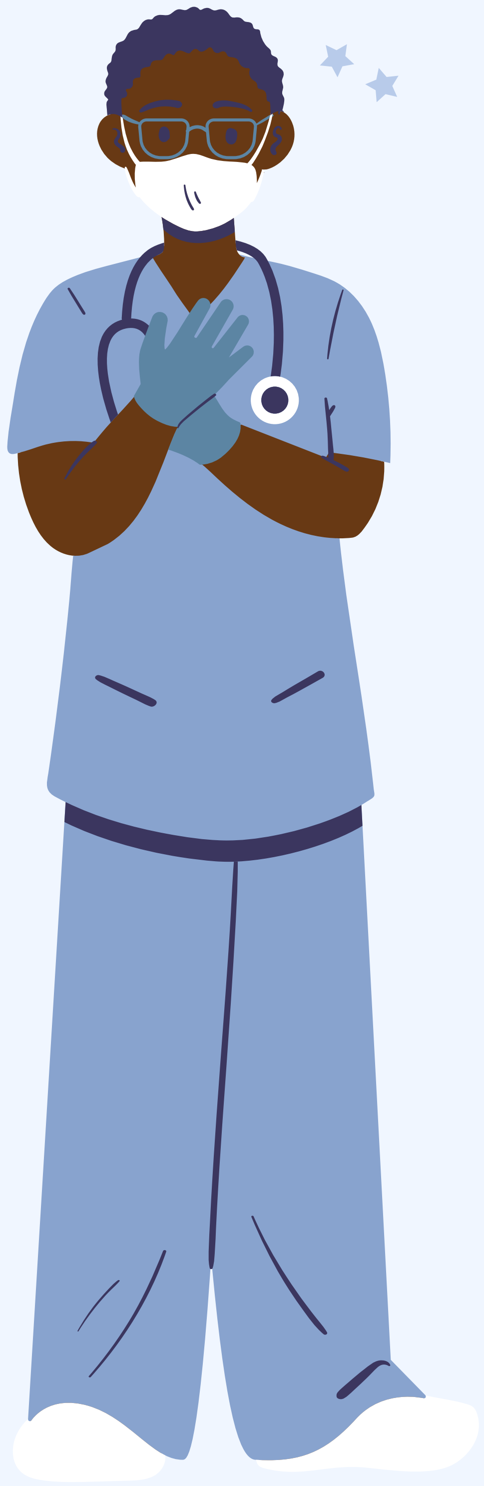
Capstone Project - COVID-19 Analysis and Forecasting

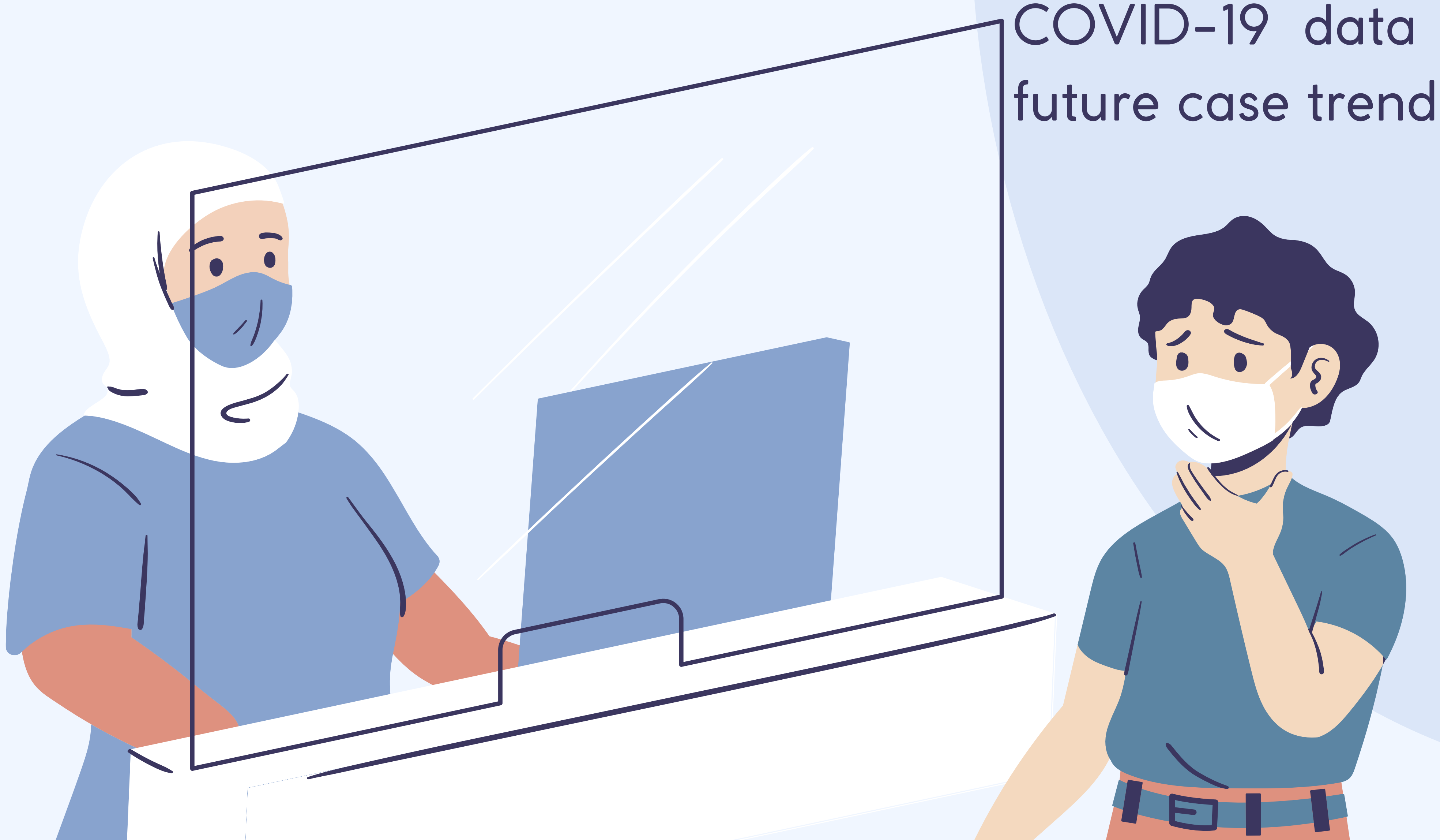
BY R LIKITH

Table of Contents



1. Problem Statement

The COVID-19 pandemic has significantly impacted public health and global economies. The rapid spread of the virus created challenges in forecasting case numbers, understanding transmission patterns, and managing healthcare resources. This project aims to analyze historical COVID-19 data and develop predictive models to forecast future case trends.



2. Project Objective

The primary objectives of this project are:

- To analyze historical COVID-19 cases and death rates.
- To identify patterns and trends in infection rates over time.
- To develop a predictive model that forecasts future case trends.
- To assess the impact of external factors such as lockdowns, government policies, and vaccination rates.

3. Data Description

The dataset includes:

- Date: Daily recorded timestamps.
- Confirmed Cases: The number of positive COVID-19 cases.
- Deaths: Total fatalities due to COVID-19.
- Recovered Cases: Individuals who have recovered from the virus.
- Testing Data: Number of COVID-19 tests conducted.
- Vaccination Rates: The percentage of the population vaccinated.

4. Data Pre-processing Steps and Inspiration

The data preprocessing steps included:

- Handling Missing Values: Used forward-fill and interpolation techniques to fill gaps.

- Date Conversion: Transformed the Date column into a datetime format.

- Feature Engineering:
  - Created new columns such as Daily Case Growth Rate and Active Cases.
  - Standardized data to ensure consistency across different countries.

5. Choosing the Algorithm for the Project

For time series forecasting, we explored the following models:

- Facebook Prophet Model: Handles seasonality, trends, and missing data effectively.
- LSTM (Long Short-Term Memory): A deep learning model used for time series forecasting.

6. Motivation and Reasons for Choosing the Algorithm

- Prophet Model: Works well with irregular time series and missing values.
- LSTM: Captures long-term dependencies in sequential data and provides accurate forecasts.
- Comparing both models helps in selecting the best approach for real-world predictions.

7. Assumptions

- The data is reliable and accurately reflects real-world COVID-19 trends.
- Future trends will follow historical patterns, assuming no drastic policy changes.
- Vaccination rates and government interventions significantly impact infection trends.

Model Evaluation and Techniques

The following techniques were used for model evaluation:

- Exploratory Data Analysis (EDA): Visualizing trends in case numbers and fatalities.
- Correlation Analysis: Understanding relationships between infection rates, deaths, and policy measures.

1. Predictive Modeling:
    - Implemented Prophet and LSTM models for forecasting.
    - Evaluated model performance using Mean Absolute Percentage Error (MAPE).

9. Inferences from the Analysis

Key insights from the analysis:

- COVID-19 cases exhibit strong seasonality, with peaks occurring at regular intervals.
- Lockdowns and vaccination drives significantly reduced infection rates.
- LSTM outperformed Prophet in long-term forecasting due to its ability to capture nonlinear patterns.

10. Future Possibilities of the Project

- **Incorporate external factors such as mobility data and weather conditions.**

Overview 01

Enhance deep learning models using hybrid architectures like CNN-LSTM.

Overview 02

Develop a real-time COVID-19 prediction dashboard for healthcare planning.

Conclusion

This project successfully analyzed and forecasted **COVID-19** trends using machine learning models. The insights derived will help in public health decision-making and resource allocation. Future improvements can enhance prediction accuracy and broaden the scope of analysis.

References :

- COVID-19 Open Data Sources
- Facebook Prophet Documentation
- Deep Learning for Time Series Forecasting

Thank you for your attention

Do you have any questions ?

www.linkedin.com/in/rlikithkumar