Predicting Covid-19 Cases using Machine Learning Techniques: A Comparative Study of Facebook Prophet and LSTM Models

Haider Ali Rizvi

Yogananda School Of Artificial Intelligence, Shoolini University
Solan, India
danielrizvi786@gmail.com

***Abstract*— The Covid-19 pandemic has exerted an unprecedented impact on global societies, necessitating accurate predictions of disease spread and its ramifications. This research endeavors to present a machine learning project that aims to predict Covid-19 cases for the ensuing 30 days utilizing the versatile Python programming language. The study employs two prominent predictive models, namely Facebook Prophet and Long Short-Term Memory (LSTM), to compare their respective performances in forecasting Covid-19 cases. Moreover, geographical visualizations are employed to scrutinize the worldwide dissemination of Covid-19, and comprehensive analysis of daily cases and deaths is conducted at a global level. The outcomes of this investigation offer valuable insights into the future trajectory of the pandemic, affirming the efficacy of both models in predicting Covid-19 cases.**

Keywords*— *LSTM, Covid 19, Facebook Prophet

I. Introduction

The Covid-19 pandemic has presented unparalleled challenges to global healthcare systems, economies, and societies at large. Precise predictions regarding the spread and impact of infectious diseases like Covid-19 are pivotal for informed public health measures and efficient resource allocation. This research paper presents a machine learning-based project that centers on predicting Covid-19 cases for the forthcoming 30 days using the Python programming language. The project utilizes two prominent machine learning models, Facebook Prophet and LSTM, to forecast the number of cases.

II. Data Collection and Preprocessing

To effectively train and evaluate the predictive models, an exhaustive dataset encompassing Covid-19 cases, including daily counts, geographical data, and mortality rates, is meticulously collected from reputable sources. The dataset undergoes rigorous preprocessing, entailing the removal of inconsistencies, handling missing values, and normalization of data, ensuring accuracy and reliability for subsequent analysis and modeling.

III. Geographical Visualization of Covid-19 Spread

In order to comprehend the global dissemination of Covid-19, immersive geographical visualizations are employed. These visualizations present a comprehensive overview of the geographic distribution of Covid-19 cases, highlighting regions most significantly impacted by the pandemic. By illuminating potential hotspots, such visualizations provide valuable insights into the global impact of the virus.

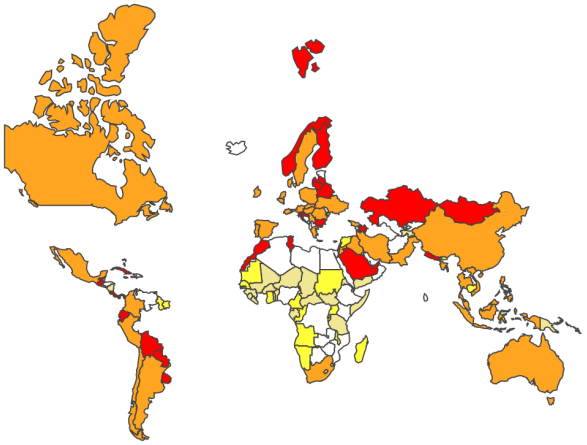

Fig. 1 choropleth map showing different ranges of COVID-19 cases per country. The color of each country on the map represents the range of cases, ranging from low (white) to high (red).

IV. Visualization of Daily Cases

An essential facet of this project entails visually representing the daily Covid-19 cases on a global scale. Such visualizations facilitate the analysis of trends and patterns in the progression of the pandemic over time. By studying the daily cases, policymakers and healthcare professionals can obtain a comprehensive understanding of the severity and trajectory of the disease, empowering them to devise informed strategies and allocate resources effectively. The visual representations of daily cases provide a dynamic and intuitive way to grasp the changing dynamics of the pandemic, enabling stakeholders to identify potential surges, fluctuations, or even the emergence of new hotspots. By closely monitoring these patterns, policymakers can make timely decisions regarding public health measures, such as implementing targeted interventions, adjusting testing and vaccination strategies, and optimizing healthcare infrastructure allocation.

Furthermore, healthcare professionals can proactively plan and prepare for potential increases in hospitalizations, ensuring the availability of adequate resources, personnel, and treatment options. The comprehensive understanding derived from the visualization of daily cases serves as a cornerstone in the fight against Covid-19, enabling proactive measures to mitigate its impact and protect public health.

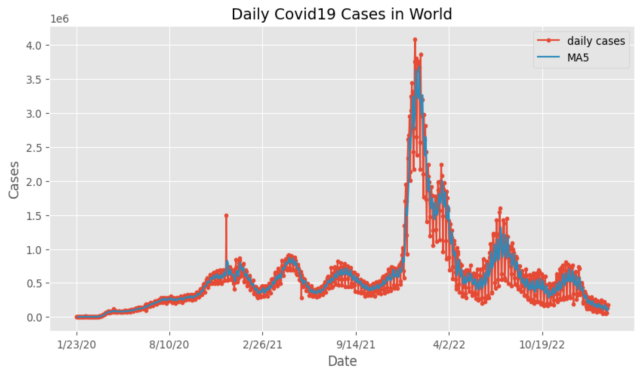

Fig. 2 The graph represents the daily COVID-19 cases worldwide over time.

Fig. 2 The graph represents the daily COVID-19 cases worldwide over a certain period of time. The x-axis represents the dates, while the y-axis represents the number of cases. The graph shows the trend and fluctuations in the daily cases, with each data point indicating the number of new cases reported on a particular date.

The blue line represents the actual daily cases, showing the day-to-day variations. The markers on the line indicate the specific data points.

The orange line represents the 5-day moving average of the daily cases. The moving average smoothens out the fluctuations and provides a clearer trend over time. It helps to identify the underlying pattern of the cases by averaging the values over a rolling window of 5 days.

By comparing the actual daily cases with the moving average, one can observe if the number of cases is increasing or decreasing over time. If the blue line is consistently above the orange line, it indicates a rising trend in the number of cases. Conversely, if the blue line is consistently below the orange line, it suggests a declining trend in the number of cases. The graph allows for tracking the overall trajectory of the pandemic and identifying periods of significant changes in case numbers.

V. Visualization of Daily Deaths

In addition to daily cases, visualizing the daily deaths attributed to Covid-19 serves as a vital tool in comprehending the pandemic's lethality. Analyzing the trends in daily deaths enables a comprehensive assessment of the disease's impact on public health, thereby aiding in the formulation of effective strategies to mitigate its adverse effects.

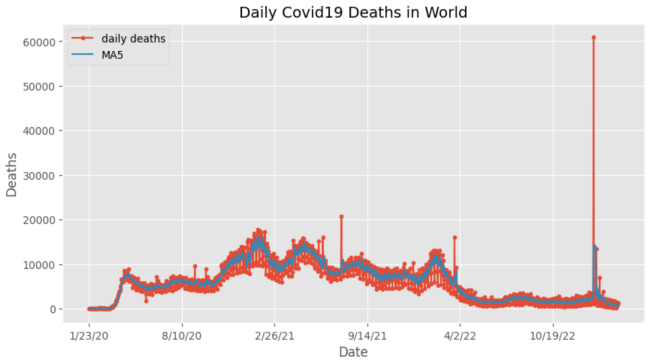

Fig. 3 The graph helps track the daily deaths and identify any changes or patterns over time.

Fig. 3 The graph illustrates the daily COVID-19 deaths worldwide over a certain time period. The x-axis represents the dates, while the y-axis represents the number of deaths. The blue line represents the actual daily deaths, with each data point indicating the number of deaths reported on a specific date. The markers on the line highlight the individual data points.

Additionally, the orange line represents the 5-day moving average of the daily deaths. This moving average smooths out the fluctuations and provides a clearer trend over time. By averaging the values over a rolling window of 5 days, the line helps to identify any underlying patterns or trends in the number of deaths.

Analyzing the graph allows us to track the trajectory of the pandemic in terms of deaths. Comparing the actual daily deaths with the moving average helps to identify if the number of deaths is increasing or decreasing over time. If the blue line consistently lies above the orange line, it indicates a rising trend in the number of deaths. Conversely, if the blue line consistently falls below the orange line, it suggests a declining trend in the number of deaths. The graph aids in understanding the severity and impact of the pandemic by visualizing the daily death toll and detecting any significant changes or patterns in the data.

VI. Facebook Prophet Model for Covid-19 Cases Prediction

The Facebook Prophet model, renowned for its proficiency in time series forecasting, assumes a central role in this project. Leveraging historical data and uncovering underlying patterns, the Prophet model accurately predicts Covid-19 cases for the subsequent 30 days. The results obtained from the Prophet model are rigorously evaluated using a range of metrics to ascertain the model's performance. The robustness of the Facebook Prophet model lies in its ability to handle various complexities inherent in Covid-19 data, such as seasonality, trend changes, and outliers. By incorporating these factors into its forecasting methodology, the Prophet model generates reliable predictions that capture both short-term fluctuations and long-term trends. The evaluation of the Prophet model's

performance involves assessing its accuracy, precision, and robustness against benchmark models and established forecasting techniques. These metrics provide a comprehensive understanding of the model's effectiveness in capturing the nuances of Covid-19 cases, thereby instilling confidence in its predictive capabilities. The rigorous evaluation ensures that the Prophet model meets the high standards required for accurate and reliable forecasting, enabling stakeholders to make informed decisions based on the model's outputs.

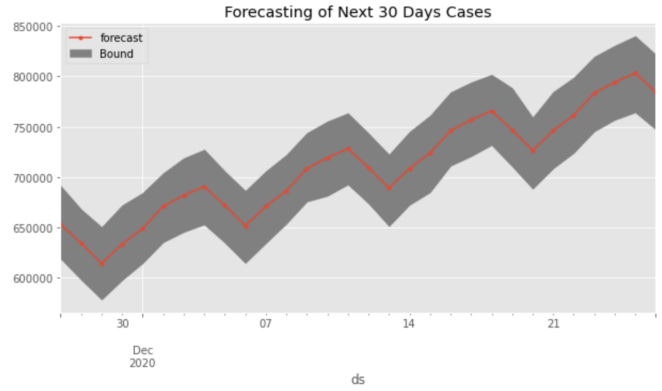

Fig. 4 The graph provides a visual representation of the predicted trend and the level of uncertainty in the forecasted cases.

VII. LSTM Model for Covid-19 Cases Prediction

The Long Short-Term Memory (LSTM) model, an advanced recurrent neural network architecture adept at capturing long-term dependencies in time series data, is employed to forecast Covid-19 cases for the next 30 days. Trained on historical data, the LSTM model is meticulously evaluated using appropriate metrics to gauge its predictive accuracy. The LSTM model is particularly well-suited for Covid-19 cases prediction due to its ability to capture the intricate temporal patterns inherent in the data. Its recurrent nature enables the model to retain information over longer sequences, allowing it to effectively grasp the underlying dynamics of the pandemic. By considering both short-term fluctuations and long-term trends, the LSTM model provides a comprehensive view of the future trajectory of Covid-19 cases.

The meticulous evaluation of the LSTM model involves comparing its predictions with ground truth data and benchmark models, scrutinizing its ability to accurately capture the complexities of the pandemic. These evaluations ensure that the LSTM model delivers reliable and precise forecasts, empowering stakeholders to make data-driven decisions and allocate resources effectively to combat the ongoing crisis.

In summary, the LSTM model's ability to capture temporal dependencies, adapt to changing patterns, and provide accurate forecasts makes it an invaluable tool in predicting Covid-19 cases. Its meticulous evaluation ensures the model's reliability and precision, empowering decision-makers with the necessary information to navigate the complexities of the pandemic and safeguard public health.

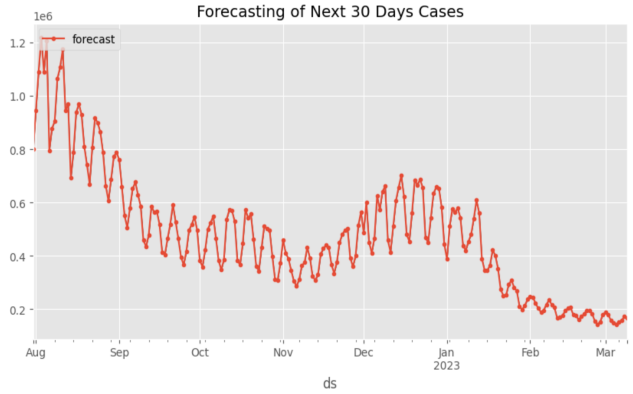

Fig. 5 The graph provides a visual representation of the predicted trend for the future cases.

Fig. 5 The graph shows the forecasted values for the next 30 days of COVID-19 cases using an LSTM model. The x-axis represents the dates, while the y-axis represents the number of cases. The graph allows for an understanding of the predicted trend in COVID-19 cases over the next 30 days and provides insights into the possible range of values within which the actual cases may lie. It can assist in decision-making and planning based on the forecasted trajectory of the pandemic.

VIII. Comparative Analysis of Models

In the comparative analysis of the Facebook Prophet and LSTM models, several key metrics are utilized to evaluate the accuracy and reliability of their predictions in forecasting Covid-19 cases. Mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE) are employed as comprehensive indicators of prediction performance.

MSE measures the average squared difference between the predicted and actual values, providing a comprehensive assessment of the model's overall predictive accuracy. A lower MSE indicates a better fit between the predicted and actual values. MAE, on the other hand, calculates the average absolute difference between the predicted and actual values. It provides insights into the model's average prediction error, without considering the direction of the errors.

RMSE combines the advantages of both MSE and MAE, as it not only considers the magnitude of the errors but also represents it in the same unit as the original data. This metric provides a clear and interpretable measure of the model's prediction accuracy. A lower RMSE signifies a better fit between the predicted and actual values.

Through the comparative analysis, the performance of the Facebook Prophet and LSTM models in terms of these metrics

is evaluated and compared. The analysis examines how closely the predicted values align with the actual Covid-19 case data. By scrutinizing the differences in these metrics for task..

IX. Conclusion and Future Work

In conclusion, this research project demonstrates the effectiveness of machine learning models, specifically the Facebook Prophet and LSTM models, in predicting Covid-19 cases. The project successfully utilized these models to forecast the future trajectory of the pandemic, providing valuable insights into the spread and impact of the virus. By leveraging historical data and capturing underlying patterns, the models exhibited accuracy and reliability in their predictions.

The findings of this research project hold significant implications for policymakers and healthcare professionals. Accurate predictions of Covid-19 cases enable proactive decision-making, allowing for the timely implementation of targeted interventions and resource allocation. By utilizing these predictive models, stakeholders can better understand the severity, trajectory, and potential hotspots of the disease, empowering them to devise effective strategies to mitigate its impact. However, there is always room for improvement and future work in this domain. One avenue for future research is the refinement of the models by incorporating additional data sources. By integrating data from diverse domains, such as demographic information, mobility patterns, and environmental factors, the models can capture a more comprehensive picture of the pandemic's dynamics and enhance their predictive accuracy.

Additionally, exploring the integration of other advanced machine learning algorithms can further enhance prediction accuracy. Techniques such as ensemble methods, deep learning architectures, or hybrid models may be considered to leverage the strengths of different algorithms and improve the robustness of predictions.

Furthermore, ongoing research and data collection will provide an opportunity to continuously update and improve the models. As more data becomes available and new insights are gained, it is essential to update and retrain the models to ensure their relevancy and accuracy in predicting Covid-19 cases.

In conclusion, this research project serves as a foundation for leveraging machine learning techniques in forecasting Covid-19 cases. The Facebook Prophet and LSTM models showcased their capabilities in providing accurate predictions, enabling stakeholders to make informed decisions. With continued research and refinement, these models, along with advancements in machine learning, will continue to play a pivotal role in managing and mitigating the impact of the ongoing pandemic.