ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940
www.ijsem.org
Vol 09, Issuse.4 Dec 2018

A STUDY ON THE DIGITAL PAYMENT SERVICES IN INDIA

Sandhya, M.Sc , Ph.D,*1, Santhosh MCA, *2, sreedhar MCA, *3.

ABSTRACT

Since 2019, the global impact of COVID-19, a severe respiratory illness, has been extensive, resulting in widespread fatalities and posing substantial threats to global economies, political landscapes, and society at large. In light of this critical scenario, accurately predicting the future trajectory of COVID-19 becomes imperative. This comparative study rigorously investigates the efficacy of four distinct time-series analysis models—the ARIMA model, the Prophet model, the Long Short-Term Memory (LSTM) model, and the Transformer model—in forecasting the forthcoming case trends of COVID-19 across six countries. The research highlights the remarkable accuracy of the LSTM model, positioning it as a pivotal tool for anticipating the spread of COVID-19. Leveraging such predictive capabilities can empower nations to bolster preparedness and awareness, aiding in more effective control of the virus's dissemination.

This project serves as a significant contribution to the on-going efforts to understand and mitigate the impact of COVID-19. By identifying the LSTM model's superior performance in predicting case trends, the research underscores the potential of advanced time-series analysis methodologies in informing public health strategies and guiding proactive measures to curtail the virus's spread. The insights gained from this study can be instrumental in refining predictive models, fostering international collaboration, and ultimately mitigating the devastating effects of the on-going global health crisis.

Key words: Time-series analysis, ARIMA, Prophet

I. A Comparative Analysis: Evaluating Time-Series Analysis Approaches to Predict the Trend of COVID-19 Cases

II . INTRODUCTION

This chapter provides an overview of the thesis project, encompassing the background of COVID-19, current time-series analysis methods, the research problem, goals, methodology, delimitations, and the thesis structure. Backgroud

In December 2019, the emergence of COVID-19, caused by the SARS-CoV-2 virus, led to a global pandemic with over 214 million confirmed cases and four million deaths to date. The highly contagious disease has prompted widespread lifestyle changes and economic impacts, necessitating diverse preventive measures. Time-series data analysis methods, including statistical approaches like ARIMA and machine learning techniques such as LSTM, have proven effective in understanding and predicting the transmission dynamics of infectious diseases like COVID-19. Analyzing COVID-19 poses unique challenges due to its severity, rapid spread, and testing difficulties. Key issues include determining the relevant statistics for trend prediction, adapting methods to accommodate data growth, and establishing evaluation criteria for model performance. The purpose is to comprehend the disease pattern, enabling effective measures to curb its spread, with ethical considerations emphasizing the importance of continuous research and acknowledging potential inaccuracies in predictions. This sustainable project, using minimal resources, contributes positively by identifying the best-performing time-series model for COVID-19 prediction, serving as a foundation for further research and promoting credible forecasting.

*1. Lecturer in Mathematics, Siva Sivani Degree College, NH-44 Kompally, Secunderabad-500100, Telangana, India.

*2, Lecturer in Computer science Siva Sivani Degree College, NH-44 Kompally, Secunderabad-500100, Telangana, India.

*3. Lecturer in Computer Science, Siva Sivani Degree College, NH-44 Kompally, Secunderabad-500100, Telangana, India

ISSN2454-9940
www.ijsem.org
Vol 09, Issuse.4 Dec 2018

Core Issue:
Despite the success of time-series analysis methods in examining infectious diseases, analysing COVID-19 poses distinct challenges due to its severity, making it one of the deadliest health crises in recent times with far-reaching impacts globally. The sheer volume of total cases, heightened contagiousness, rapid transmission, and the intricacies of accurate testing and patient tracking compound the complexity of COVID-19 analysis compared to milder diseases. The central concern lies in assessing the continued relevance and effectiveness of existing time-series analysis methods within the unique context of COVID-19, and determining the optimal method, if one exists, for accurately depicting its disease trend and predicting future trajectories .
Purpose:
Understanding the underlying patterns in the disease curve aids in gaining insights into the nature of the disease and assessing the current severity of the pandemic. This knowledge is essential for implementing targeted measures to curb the spread while minimizing associated losses. Given the global significance of COVID-19, a comprehensive understanding of its curve and the ability to predict future trends are imperative.

o Ethics and Sustainability:

Accurate prediction of COVID-19 holds significant benefits, including saving lives, minimizing losses, implementing effective measures, and aiding in global economic recovery. However, the ethical challenge lies in the inherent uncertainty of predictions. Acknowledging the potential vagueness or inaccuracy of predictions is crucial, as misleading information could worsen situations and lead to greater losses. To address this, continuous research on COVID-19 prediction is essential beyond this project, and this work should not be the sole basis for imposing behaviours related to COVID-19. Despite this ethical dilemma, the project's value lies in identifying the best-performing time-series model, serving as a foundation for further research and model development, thereby enhancing prediction credibility.

The project exhibits positive sustainability impact, conducted once with minimal resource consumption during experiments and no significant pollutants emitted. All data and modelling utilities are sourced from online platforms or open-source repositories. The project's results contribute positively to society, and there is no sustainability concerns associated with its execution.

Related Work:
In response to the COVID-19 pandemic, extensive research has focused on analysing the virus's trend and making predictions using diverse methodologies. Numerous studies have undertaken a comparative analysis of various time-series models, offering insights into their effectiveness. Several notable contributions in this domain are summarized below:

- K.E. ArunKumar et al. [23] conducted a 60-day forecast of COVID-19, employing deep layer Recurrent Neural Networks (RNNs). Their research encompassed 10 countries with the highest confirmed cases, introducing customized RNN models tailored to each country. The study compared Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) units, finding that LSTM did not consistently outperform GRU, with GRU excelling in certain countries. The conclusion emphasized the necessity of developing a DL model utilizing confirmed, recovered, and death cases to achieve higher prediction accuracy.
- Khondoker Nazmoon Nabi et al. [24] focused on the COVID-19 outbreak in Brazil, Russia, and the United Kingdom, employing four DL models: LSTM, GRU, CNN, and Multivariate Convolutional Neural Network (MCNN). The study highlighted CNN's superior accuracy and consistency compared to other methods. LSTM's limitations in capturing trends, particularly in the absence of seasonality, were noted, with insufficient data observed for LSTM to effectively capture data characteristics.
- Christophorus Beneditto Aditya Satrio et al. [25] conducted a comparative study of ARIMA and Prophet for forecasting COVID-19 spread in Indonesia. Utilizing confirmed, death, and recovered case data, the study concluded that Prophet exhibited superior accuracy over ARIMA, achieving a 91% accuracy for the entire confirmed case prediction. However, the research acknowledged limitations in training data sufficiency.
- Ayodele Ariyo Adebiyi et al. conducted a comparison between ARIMA and Artificial Neural Network (ANN) models for predicting stock prices. Using closing stock prices as daily variables, their research included 5680 observations and demonstrated that both models achieved good performance. However, ANN exhibited higher forecast accuracy, establishing its superiority over ARIMA in stock price prediction. The study suggested future exploration of hybrid models to further enhance performance.

A. Time-Series Analysis :

Time-series data, organized chronologically, feature consistent intervals and find applications in statistics, economics, finance, and forecasting. Time-series analysis involves analytical methods to discern temporal components, unveiling meaningful patterns and statistics. Utilizing time-series models, these analyses fit the data to reveal its characteristics. Time-series forecasting predicts future values based on historical observations. This project explores four time-series analysis methods for predicting COVID-19 trends, each chosen for its proven expertise in specific domains. Despite their historical effectiveness, their adaptability and performance in disease trend prediction remain uncertain. The selected methods—ARIMA, Prophet, LSTM, and Transformer—will be introduced and evaluated to determine their applicability and performance in forecasting COVID-19 trends. The study aims to leverage the strengths of these established methods in their respective domains to enhance predictions in the dynamic context of COVID-19.

First present a brief introduction of the time-series analysis methods used in this project, i.e., ARIMA, Prophet, LSTM:-

i) ARIMA

ARIMA (AutoRegressive Integrated Moving Average) is a widely used time-series model in statistics and econometrics, serving as a generalization of the Autoregressive Moving Average (ARMA) model. Overcoming the ARMA's

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijsem.org
Vol 09, Issuse.4 Dec 2018

limitation to stationary time-series data, ARIMA incorporates differencing into the model. As a form of regression analysis, it is extensively applied for data understanding and future predictions based on historical values. Non-seasonal ARIMA is denoted as ARIMA(p, d, q), where p represents the lag order in the Autoregressive component, d is the degree of differencing for stationarity, and q signifies the order of the Moving Average component.

For Seasonal Autoregressive Integrated Moving Average (SARIMA) models, denoted as SARIMA (p, d, q) (P, D, Q, m), additional seasonal terms (P, D, Q) are introduced, with m representing the number of periods in each season. Applications of ARIMA are diverse; for instance, Javier Contreras et al. employed ARIMA models to analyse hourly prices in Spanish electricity markets, achieving a reasonable average error of approximately 10% in predicting next-day prices. Rajesh G. Kavas Seri and Kithira Seetharaman utilized f-ARIMA for wind speed forecasting, demonstrating a 42% improvement in accuracy compared to existing methods. ARIMA has also been applied in disease prediction, showcasing its versatility in various domains.

ARIMA is a time-series model in statistics and econometrics, a generalization of the Autoregressive Moving Average (ARMA). It incorporates differencing to overcome the ARMA's limitation to stationary time-series data. ARIMA is widely used for data understanding and future predictions based on historical values. Non-seasonal ARIMA is denoted as ARIMA(p, d, q), while Seasonal Autoregressive Integrated Moving Average (SARIMA) models introduce additional seasonal terms. ARIMA has diverse applications, such as analyzing hourly prices in Spanish electricity markets, improving wind speed forecasting accuracy by 42%, and predicting disease. Applications include predicting next-day prices with an average error of approximately 10%, wind speed forecasting with a 42% improvement, and disease prediction.

ii) Prophet

Prophet, developed by Facebook's Core Data Science team in 2017, is an open-source forecasting procedure available in Python and R. Described in the paper "Forecasting at Scale," Prophet employs an additive model capturing non-linear trends with yearly, weekly, and daily seasonality, including holiday effects. The formulation,

$Y(t) = g(t) + s(t) + h(t) + t,$

Breaks down into the predicted value (y (t)), trend function (g (t)), seasonal changes (s (t)), holiday effects (h (t)), and the normally distributed error (t). Prophet excels in handling strong seasonal effects, seamlessly managing missing data or shifts in the series. With fast and accurate performance, Prophet offers user-friendly parameters for model customization, as depicted in the example plot showing observed values, predicted values (dark blue), and uncertainty interval bounds (light blue). Prophet's high forecasting accuracy has made it widely versatile across diverse domains. Notably, Işil Yenidoğan et al. employed Prophet for Bitcoin forecasting, achieving a remarkable R2 value of 94.5% in a 90-day forecast based on training data from May 3rd, 2016, to August 30th, 2018. In hydrology, Georgia A. Papacharalampous1 and Hristos Tyralis utilized Prophet for daily stream flow forecasting, demonstrating comparable accuracy to random forests. Cong Xian et al. applied Prophet to analyse and predict daily reported cases of hand, foot, and mouth disease (HFMD), successfully detecting strong seasonality and holiday effects crucial for disease precautions. Prophet's success extends to finance, climate, medical, and various research fields, showcasing its effectiveness in diverse applications.

The model is known for its ability to handle strong seasonal effects and manage missing data or shifts in the series. It offers fast and accurate performance, with user-friendly parameters for model customization. Prophet's high forecasting accuracy has made it versatile across diverse domains, such as finance, climate, medical, and research fields. Notable applications include Bitcoin forecasting, hydrology, and hand, foot, and mouth disease analysis and prediction. The model's effectiveness in these fields is evident in its ability to handle strong seasonality and holiday effects.

iii) LSTM

LSTM (Long Short-Term Memory) was introduced by Sapp Hoch Reiter and Jürgen Schmidhuber as an efficient Recurrent Neural Network (RNN) architecture, addressing the vanishing gradient problem in traditional RNNs during training. Widely applied in tasks like handwriting and speech recognition, LSTM excels in processing time-series data, capturing hidden information between events with unknown intervals. A typical LSTM unit includes a cell, input gate, output gate, and forget gate, allowing it to manage short-term and long-term memory. The structure involves activation vectors and gates that cooperatively function to generate new memory and output. LSTM's versatility is evident in applications such as sequence tagging, where bidirectional LSTM achieves state-of-the-art accuracy with less dependency on word embedding, and in precipitation prediction, where convolutional LSTM outperforms other methods. Original LSTM models also exhibit high performance, as seen in residential load forecasting, where vanilla LSTM outperforms alternative methods in dataset comparisons.

Long Short-Term Memory (LSTM), introduced by Sepp Hochreiter and Jürgen Schmidhuber [16], revolutionized Recurrent Neural Network (RNN) architecture in Deep Learning (DL). Specifically designed to mitigate the vanishing gradient problem inherent in traditional RNNs during training, LSTMs find application in diverse tasks like handwriting and speech recognition. Particularly well-suited for processing time-series data, LSTMs excel at capturing latent information between events with unknown intervals.

A standard LSTM unit comprises a cell, an input gate, an output gate, and a forget gate, each playing a crucial role. Interactions among these components facilitate the acceptance of short-term memory, long-term memory, and input, generating new short-

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijsem.org
Vol 09, Issuse.4 Dec 2018

term memory, long-term memory, and output. The LSTM unit structure, illustrated in the figure, involves activation vectors (ft, it, ot) for the forget, input, and output gates, respectively. Input vector xt and hidden state vector ht contribute to the LSTM unit's functionality. Additionally, ct and c˜t denote the cell state vector and cell input activation vector, respectively.

The cooperative functioning of gates is expressed through mathematical formulations, involving activation functions (σg, σc, σh) and element-wise product operations (◦). The weight matrices (W, U) and bias vector (b) further contribute to the LSTM's intricate architecture.

Over two decades of study, LSTMs and their variations have found widespread application. Researchers, such as Zhiheng Huang et al. [17] employing bidirectional LSTMs for sequence tagging, have achieved state-of-the-art accuracy with reduced dependence on word embedding. In novel approaches, Xingjian Shi et al. [18] utilized convolutional LSTMs for precipitation prediction, surpassing existing methods. Even in its original form, LSTMs, as demonstrated by Weicong Kong et al., exhibited remarkable performance in residential load forecasting, outperforming alternative methods in the dataset.

iv) Transformer

Introduced in 2017, the Transformer model has swiftly garnered widespread acclaim as a cutting-edge deep learning (DL) architecture, revolutionizing sequential data processing through the implementation of attention mechanisms. Distinguishing itself from Recurrent Neural Networks (RNN), the Transformer model diverges by relinquishing reliance on the sequential order of data. Instead, it leverages attention mechanisms to furnish contextual understanding for any position within the sequence. This distinctive characteristic empowers the Transformer with parallelization capabilities, translating to significantly reduced training times. Notably, the Transformer model has risen to prominence as the architecture of choice, particularly in domains such as Natural Language Processing (NLP) and Computer Vision (CV), eclipsing traditional RNN models like LSTM owing to its unparalleled efficacy and adaptability.

The transformative impact of the Transformer model lies in its departure from sequential constraints, offering a paradigm shift in data processing strategies. Unlike RNNs, which grapple with sequential dependencies, the Transformer excels by providing context through attention mechanisms without being tethered to a specific order. This innovative approach not only enhances efficiency but also facilitates widespread adoption, positioning the Transformer as a frontrunner in state-of-the-art DL architectures. Its success in NLP and CV underscores its versatility and applicability, solidifying its standing as a revolutionary model that transcends the limitations of traditional sequential data processing.

The Transformer model consists of an encoder stack on the left and a decoder stack on the right, featuring a multi-head self-attention layer and a feed-forward layer in the encoder. The decoder includes a masked multi-head self-attention layer, a multi-head self-attention layer, and a feed-forward layer, reducing path length for improved learning of long-range dependencies.
This design contributes to more interpretable models. Despite its success in Natural Language Processing (NLP) and Computer Vision (CV), the Transformer has been adapted for time-series forecasting. Sifan Wu et al. introduced the Adversarial Sparse Transformer, demonstrating its effectiveness for short-term and long-term prediction through extensive real-world dataset experiments. Neo Wu et al. applied Transformer-based models to forecast influenza-like illness (ILI) with results favorably comparable to state-of-the-art methods.

III. METHODS

Data Collection

The dataset utilized in this project is sourced from the Johns Hopkins University Centre for Systems Science and Engineering (JHU CSSE) repository. Updated daily since January 22nd, 2020, this dataset aggregates information from various sources, providing accurate and comprehensive statistics on COVID-19 case reports in the United States and numerous other countries. The dataset includes a time-series summary of COVID-19 cases, aligning well with the research objectives of this project. The forthcoming sections will delve into the data selection and pre-processing methodologies.

- Data Source:- As is mentioned above, we choose the time-series data from the COVID-19/ csse_covid_19_data/csse_covid_19_time_series/ folder as our data source.∗ Three files are downloaded, i.e., the global confirmed cases, the global death cases, and the global recovered cases. The file names are time_series_covid19_confirmed_global.csv, etc. For the simplicity of experiments, we first use the global confirmed case as a single data source, which means we are considering a univariate scenario.
- Data Pre-Processing:- Due to the significant number of cases, the decision was made to use daily new cases as the input data (or training data) for all time-series analysis models. This is expressed by the equation
  $N(n) = C(n) - C(n - 1)$

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940
www.ijsem.org
Vol 09, Issuse.4 Dec 2018

where N(n) represents the new cases on the nth day, and C(n) represents the cumulative case number on the nth day. By calculating the daily cases for each country and replacing Turkey with the UK for better comparative analysis, a clearer understanding of case trends is achieved, as depicted in Figure 3.5. Consequently, each country now has a record of daily new cases over 500 days.

Notably, negative values were observed, particularly in France (evident in Figure 3.5d) and minor outliers in the UK. To address this issue, negative values were removed from the data, and the missing values were replaced with the data from the previous day. This approach, from a statistical standpoint, serves to eliminate errors that may confuse the models and ensures a smoother and more reasonable curve for time-series analysis.

IV. EXPERIMENT MODELS AND PROCEDURE IN OUR EXPERIMENTS

Trial Models and Procedure In our trials, 5 different time- series analysis styles models(including a naïve birth model) will be used to compare with each other in prognosticating unborn trend of COVID- 19 cases. They're a naïve ensemble model (birth model), ARIMA, Prophet, LSTM, and Transformer, independently. Some redundant detail of each model is explained as follows.

• Birth model the birth model is a naïve combination of simple standard models for univariate series. In the trials, we implement this model as the normal of some standard models, i.e., naïve drift model, naïve mean model, and naïve seasonal model. The delineations of these models can be set up in the operation Programming Interface (API) reference of Darts Python package.

• ARIMA In order to use ARIMA model, parameters, i.e., p, d, and q explained in Section2.1.1, must be determined beforehand. This can be done by checking the stationary and autocorrelation structure using Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF), or by simply using AutoARIMA model in some packages that can identify the data characteristics and automatically set the parameters.

• Prophet to help possible negative prognosticated values, some measures need to be performed. We first take the natural logarithm of data as input training data) to feed into the model, and also take natural exponential of the affair (prognosticated values) as the result. Negative perpetuity is treated as NaN to avoid fine problems.

• LSTM Hyper parameters need tuning. This includes unit number, learning rate, etc. generally we can use grid hunt to perform hyper parameter tuning, but some packages also give accessible tuner to accelerate this process. We'll use the ultimate approach in our trials.

i) Data Processing:

The case data, whether normalized or not, is treated as univariate data. To capture relevant information, we introduce a look-back approach, setting the number of relevant days to 10. This implies that the case number on a given day is considered relevant to the case numbers in the previous 10 days. For new cases on the 1st to 10th day, we pad the sequence with 0's for unavailable days. For instance, new cases on the 2nd to 11th day contribute to predicting the new cases on the 12th day. This process results in two datasets from the original data: one serving as the feature matrix and the other as the target. These datasets are utilized for training and validation in the experiments.

ii) Parameter Tuning:

For RNNs like LSTM, fine-tuning numerous parameters and hyper parameters is crucial. The keras-tuner library in Keras accelerates this process. In our experiments, considering a single variable and a relatively simple relationship, we opt for a straightforward model structure. The model comprises one LSTM layer and one regular densely-connected Neural Network (NN) layer. Two parameters are tenable in this structure: the number of units in the LSTM layer and the learning rate. The code snippet in Listing 4.4, along with comments, elucidates the parameter tuning range.

• Motor analogous to LSTM, we need grid hunt to determine the stylish hyper parameters. It includes further parameters similar as encoder layers and decoder layers. In the coming chapter

ISSN2454-9940
www.ijsem.org
Vol 09, Issuse.4 Dec 2018

the stylish parameters will be presented for each country.

Evaluation metrics

Mean Squared Error (MSE)
In this project, the evaluation metrics employed are Mean Squared Error (MSE) and Mean Absolute Error (MAE), defined as:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2,$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|,$$

The evaluation metrics employed in this project, namely Mean Squared Error (MSE) and Mean Absolute Error (MAE), serve as robust measures to assess the accuracy of the predictive models. Calculated for the last 50 predicted values in the results, these metrics offer a comprehensive analysis of model performance during the critical forecasting period. Although alternative metrics such as R2 and F-score could potentially be introduced for a more comprehensive evaluation, the project deliberately emphasizes the simplicity and interpretability of MSE and MAE. This focused approach streamlines the assessment process, allowing for a clear comparison of model accuracies and aiding in the identification of the model that excels in predicting COVID-19 cases.

By honing in on MSE and MAE as the primary evaluation metrics, the project prioritizes clarity and ease of interpretation. These metrics not only capture the accuracy of predictions but also facilitate a straightforward comparison of different models. While acknowledging the availability of other metrics, the project's deliberate choice aligns with the goal of providing a concise and accessible evaluation framework for the diverse time-series analysis methods under scrutiny.

FUTURE WORK
While this project has addressed initial goals, several challenges and limitations have surfaced during experimentation, prompting the identification of key areas for future improvement:

1. Model Diversity: The study only incorporates four time-series analysis methods for predicting COVID-19 trends. Future work should expand the comparison to include additional models such as linear regression and Convolutional Neural Networks (CNNs) to provide a more comprehensive assessment.

2. Multivariate Analysis: The research focuses solely on univariate data. Future efforts could benefit from incorporating additional data sources, such as death and recovered cases, vaccination rates, and other COVID-19-related data. Conducting experiments in two parts—univariate prediction for all models and multivariate prediction for those supporting it (e.g., the LSTM model)—can enhance the overall experiment and optimize performance.

3. Real-Time Data Integration: The data used in experiments may become outdate. Implementing a data pipeline for direct integration of the latest data from the repository would ensure models have access to up-to-date information, facilitating more accurate analysis and future trend predictions.

4. Hardware Limitations: Current hardware constraints may hinder models from adopting more complex structures or exploring a broader parameter range. Potential solutions involve hardware upgrades or leveraging platforms like Google Cola, which should be explored in future discussions.

5. Hybrid Models: Hybrid models, combining statistical models (e.g., ARIMA) and machine learning approaches (e.g., Artificial Neural Networks), have gained prominence in time-series analysis. Investigating such models in future work could enhance predictive performance for COVID-19 trends.

V. CONCLUSIONS

This comparative study systematically evaluates the effectiveness of five distinct time-series analysis methods, encompassing a naïve baseline model, in predicting COVID-19 cases across six countries. Leveraging COVID-19 case data sourced from the Johns Hopkins University Centre for Systems Science and Engineering (JHU CSSE), we meticulously select six countries with the highest cumulative confirmed cases, strategically replacing Turkey with the United Kingdom to ensure statistical accuracy. By focusing on daily new cases and employing pre-processing techniques for data smoothing, we normalize the time-series data and subsequently apply five models—baseline, ARIMA,

ISSN2454-9940
www.ijsem.org
Vol 09, Issuse.4 Dec 2018

Prophet, LSTM, and Transformer—to fit and predict future case numbers and trends. The evaluation of prediction results using metrics like Mean Squared Error (MSE) and Mean Absolute Error (MAE) underscores the exceptional performance of the LSTM model across all scrutinized countries. With MSE consistently below 0.006 and MAE below 0.06, the LSTM model emerges as the most robust performer, surpassing other models. The Transformer model demonstrates the second-best performance, particularly excelling in three countries and exhibiting comparable accuracy to LSTM. Notably, other models, under specific circumstances, exhibit less convincing results than the baseline model, as evident in the case of India. In summation, the LSTM model proves to be a stalwart and reliable time-series analysis method, effectively capturing disease trends and furnishing accurate predictions, thereby facilitating informed measures for controlling the spread of the virus and minimizing human infections and associated losses.

This research contributes valuable insights into the field of time-series analysis for disease trends, showcasing the LSTM model as a preferred choice for accurate predictions in the context of COVID-19. The findings emphasize the importance of selecting appropriate models for forecasting, with LSTM standing out as a dependable tool for public health decision-making. By establishing LSTM's exceptional performance, this study provides a foundation for further research and model refinement, aiming to enhance the predictive capabilities of time-series analysis in the on-going battle against the global pandemic.

REFERENCES

[1] R. Allard, “Use of time-series analysis in infectious disease surveillance.” Bulletin of the World Health Organization, vol. 76, no. 4, pp. 327–333, 1998. [Online]. Available: https://www.ncbi.nlm. nih.gov/pmc/articles/PMC2305771/ [Page 2.]

[2] CSSEGISandData, “COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University,” Sep. 2021, original-date: 2020-02-04T22:03:53Z. [Online]. Available: https://github.com/CSSEGISandData/COVID-19 .

[3] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control. John Wiley & Sons, May 2015. ISBN 978-1-118-67492-5 Google-Books-ID: rNt5CgAAQBAJ.

[4]G. P. Zhang, “Time series forecasting using a hybrid ARIMA and neural network model,” Neurocomputing, vol. 50, pp. 159–175, Jan. 2003. doi: 10.1016/S0925-2312(01)00702-0. [Online]. Available: https://www. sciencedirect.com/science/article/pii/S09252 31201007020 .

[5]V. Kotu and B. Deshpande, “Chapter 12 - Time Series Forecasting,” in Data Science (Second Edition), V. Kotu and B. Deshpande, Eds. Morgan Kaufmann, Jan. 2019, pp. 395–445. ISBN 978-0-12-814761-0. [Online]. Available: https://www.sciencedirect.com/science/article/pii/ B9780128147610000125 .