
GAZİ ÜNİVERSİTESİ
Gazi
1926

FORECASTING THE SPREAD OF COVID-19 USING DEEP LEARNING AND BIG DATA ANALYTICS METHODS

Cylas KIGANDA

MASTER'S DEGREE

DEPARTMENT OF COMPUTER SCIENCE

GAZI UNIVERSITY

INSTITUTE OF INFORMATICS

JANUARY 2023

ETHICAL STATEMENT

In this thesis study, which I prepared in accordance with Gazi University Informatics Institute Thesis Writing Rules;

- I have obtained the data, information and documents I have presented in the thesis within the framework of academic and ethical rules,
- I present all information, documents, evaluations and results in accordance with scientific ethics and morals,
- I have cited all the works I have benefited from in the thesis by making appropriate references,
- I have not made any changes to the data used,
- The work I have presented in this thesis is original,

Otherwise, I declare that I accept all loss of rights that may arise against me.

Cylas KIGANDA

10/01/2023

DERİN ÖĞRENME VE BÜYÜK VERİ ANALITIĞI YÖNETMLERİ KULLANARAK COVID-19 YAYILIMININ İLERİYE DÖNÜK TAHMİNİ

(Yüksek Lisans Tezi)

Cylas KIGANDA

GAZİ ÜNİVERSİTESİ

BİLİŞİM ENSTİTÜSÜ

Ocak 2023

ÖZET

COVID-19 küresel salgınının yayılmasıyla mücadele etmek için mevcut yapay zeka araçlarının kullanımını en üst düzeye çıkaran son teknoloji önlemlere ihtiyaç vardır. Bu araştırmanın odak noktası, COVID-19 vakalarını tahmin etmek ve 5 ana Afrika bölgesindeki en savunmasız ulusları belirlemektir. Önerilen teknik, mevsimsel otoregresif entegre hareketli ortalama (ARIMA), uzun süreli bellek (LSTM) ve Prophet algoritması gibi istatistiksel ve derin öğrenme tekniklerini birleştirmiştir. Ortalama Karesel Hata, Ortalama Karesel Hata Kökü, Ortalama Mutlak Yüzde Hatası, Simetrik Ortalama Mutlak Yüzde Hatası, Pik Sinyal-Gürültü Oranı, Normalleştirilmiş Kök Ortalama Kare Hatası ve R2 skoru performans ölçümleri arasında yer almıştır. Uzun kısa süreli bellek modelinin bu araştırmada en başarılı olduğu bulunmuştur. Batı, Güney, Kuzey, Doğu ve Orta Afrika bölgelerinden Mali, Angola, Mısır, Somali ve Gabon, 61 günlük tahmin çizelgesinin sonunda COVID-19 vakalarının sayısında beklenen en büyük artışları göstermiştir. Sırasıyla yüzde 22,77, yüzde 18,97, yüzde 11,83, yüzde 10,72 ve yüzde 2,81 artışla.

Bilim Kodu	:	92431
Anahtar Kelimeler	:	Derin öğrenme, COVID-19, Yapay Sinir Ağları, uzun kısa süreli bellek, otoregresif entegre hareketli ortalama, Prophet
Sayfa Adedi	:	107
Danışman	:	Prof. Dr. M. Ali Akcayol

FORECASTING THE SPREAD OF COVID-19 USING DEEP LEARNING AND BIG DATA ANALYTICS METHODS

(M. Sc. Thesis)

Cylas KIGANDA

GAZİ UNIVERSITY

INSTITUTE OF INFORMATICS

January 2023

ABSTRACT

Cutting-edge measures that maximize the usage of available artificial intelligence tools are needed to combat the spread of the COVID-19 global epidemic. The focus of this research is to forecast COVID-19 incidences and determine the most vulnerable nations in the 5 main African regions. The suggested technique combined statistical and deep learning techniques, such as the seasonal autoregressive integrated moving average (ARIMA), long short-term memory (LSTM), and the prophet algorithm. The Mean Squared Error, Root Mean-Square Error, Mean Absolute Percentage Error, Symmetric Mean Absolute Percentage Error, Peak Signal-to-Noise Ratio, Normalized Root Mean-Square Error, and R2 score were among the performance measurements. The long short-term memory model was found to be the most successful in this research. Mali, Angola, Egypt, Somalia, and Gabon, all from the Western, Southern, Northern, Eastern, and Central African regions, had the greatest expected increases in the number of COVID-19 cases by the end of the 61-day forecast timeline, with increases of 22.77 percent, 18.97 percent, 11.83 percent, 10.72 percent, and 2.81 percent, respectively.

Science Code	:	92431
Key Words	:	Deep learning, COVID-19, Artificial Neural Networks, long short-term memory, autoregressive integrated moving average, Prophet
Page Number	:	107
Supervisor	:	Prof. Dr. M. Ali Akcayol

PREFACE AND ACKNOWLEDGEMENTS

I take this opportunity to thank my supervisor, Prof. Dr. M. Ali Akcayol for his unwavering guidance on every aspect of this thesis.

I would also wish to thank all my relatives and loved ones who have continued to support me in all situations..

CONTENTS

Page

Page

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

Figure **Page**

SYMBOLS AND ABBREVIATIONS

Symbols and abbreviations used in this study are presented below with their explanations.

Symbol	**Explanation**
%	Percent

Abbreviations	**Explanations**
ARIMA	Autoregressive integrated moving average
LSTM	Long short-term memory
MSE	Mean Squared Error
RMSE	Root Mean-Square Error
MAPE	Mean Absolute Percentage Error
SMAPE	Symmetric Mean Absolute Percentage Error
PSNR	Peak Signal-to-Noise Ratio
NRMSE	Normalized Root Mean-Square Error

1. INTRODUCTION

The coronavirus ailment (COVID-19) is a pandemic that began on December 31, 2019 in Wuhan, Hubei Province, China. At first, it was thought to be a cluster of pneumonia patients. The World Health Organization (WHO) designated COVID-19 a pandemic on March 11, 2020, after a comprehensive examination of the gravity of the spread. The SARS-CoV-2 virus causes COVID-19, which can infect anyone. In most situations, people afflicted with this disease recover without the need for aggressive therapy. Individuals may have symptoms that range from mild to moderate, according to WHO. People with persistent medical conditions, especially the aged, are much more likely to suffer serious ailments.

Whenever an infected individual coughs, sneezes, talks, or breathes, the COVID-19 virus is transmitted through small secretions that are transmitted to other humans. Once those fluids stick to places like doorknobs, the virus can be transmitted to individuals who make contact with them if they do not take the proper health measures. People should keep just one-meter distance from one another, wash their hands often or use a sanitizer, wear a safety mask, and be immunized according to WHO recommendations to combat the transmission of this virus.

There is a great need to predict the spread of the COVID-19 pandemic beforehand to allow policymakers to make appropriate planning. While doing so, it is also necessary to determine the best performing prediction model given the COVID-19 cumulative positive cases data. It is also necessary to determine the most critical nations regarding the COVID-19 spread.

Several studies have been proposed that make use of various forecasting techniques in various fields such as supply chain demand management, finance, energy, the environment, and health. Some of these studies have been described briefly in the subsections that follow.

Supply chain demand management

In the field of sales in supply chain management, forecasting is a crucial input for coordinating ahead in basic program management. Outages or excess stock, depreciation, low delivery performance, emergency demands, poor resource consumption, and volatility spillover proliferation across the upstream supply chain are all examples of poor prediction consequences as described by Nenni et al. (2013). In order to deal with the complex nature of demand prediction, there is a need for robust prediction strategies.

The demand variable is a complex parameter to model by using conventional statistical methods because it is characterized by intermittent changes, which doesn't maintain the linear distribution used by most statistical methods. This presented the need for devising improved statistical approaches such as time series bootstrapping, proposed by Willemain et al. (2004). The research made with bootstrapping statistics incorporates a probability integral change factor which is used to model the intermittency. Several other statistical approaches have been proposed, such as those based on the Poisson process model, which incorporates the Bayesian model as applied by Wang et al. (2005).

Earlier statistical methods such as the Croston's method have been found to have no consistent correlation with intermittent data by studies such as the study conducted by Shenstone and Hyndman (2005). This is because this approach has an ad hoc nature, which makes it unable to capture the stochastic properties of a robust statistical strategy.

In order to estimate sales demand in a textile clothing organization, Fumi et al. (2013) compared the Fourier transform technique with the moving average and exponential smoothing methods, utilizing prior sales information over a four-year period. The Fourier transform methodology can capture the seasonality component in a given distribution, which allows it to overcome the limitations of the other two methods. Using the mean absolute percentage error, it was discovered that the Fourier transform approach produced the best results.

To predict sales, Chen and Ou (2011) used the gray extreme learning machine, which combined the gray connection analysis with the extreme learning machine. The Taguchi technique was used in conjunction with this approach. The normalized values in this investigation ranged from -1 to 1. The Taguchi approach is a loss function, quality control, and experimental design innovation factor optimization method (Chen and Ou, 2011). This is used to track the variables that have an influence on variability.

Finance

In financial terms, exchange rates refer to the price or value of a given country's currency in relation to another country's when the two are traded. Forecasting this ratio is of major interest for several financial institutions involved in this kind of trade and also other businesses which can use the forecasted exchange rates to increase their returns after trading with foreign currency.

Political factors are some of the factors to consider when modeling exchange rate forecasts as described by Blomberg and Hess (1997). This is because political factors such as political parties, elections, and candidates have a significant effect on the economic shift in most business settings. Faust et al. (2003) investigated the effect of the quality and nature of data used for forecasting on model performance. It was discovered that performance is better when the original data is used compared to when revised or preprocessed data is used. It was also concluded in this study that real-time forecasts give better results than the actual future fundamentals.

Among the most common forecasting approaches in exchange rate forecasting is the purchasing power parity. According to Taylor and Taylor (2004), the purchasing power parity theory stipulates that, to have a unit of currency from one state with the same buying value in another state, the conventional rate of exchange must match the proportion of relative price between both nations. Panda and Narasimhan (2007) compared the performance of artificial neural networks with the random walk and the linear autoregressive models to forecast exchange rates. Other approaches include economic fundamentals, Bayesian model averaging as applied by Wright (2008), and kitchen-sink regression according to research by Li et al. (2014).

Forecasting stock returns is of major interest in the financial sector. Variables such as cash flow, price fluctuations, and perception are some of the features of most research interest when modeling stock return problems. In this field, approaches including multiple kernel learning have been proposed to integrate the link between market news archives and stock returns (X. Li et al., 2014).

In other approaches, the GARCH model has been widely applied. Herwartz (2017) applied the GARCH model with the rolling window approach to predict stock returns. On the other hand, in some studies, mixed models were employed to predict stock prices, which combined

models such as the univariate predictive regression, economic constraint, generalized non-outlier model, non-outlier positive and non-outlier momentum models as applied by Dai and Zhu (2020). This approach is observed to perform well on out-of-sample forecasts. Classical methods such as ordinary least squares are observed to have a challenge of overfitting when predicting stock returns due to the greater scope of uncertainty in stock returns as described by Petropoulos et al. (2022).

Energy

Forecasting within the energy sector has had a wide range of applications. These include aspects like energy price, energy demand, consumption, renewable energy such as wave, wind force, solar power, electrical systems forecasting and so many others. Forecasting these energy variables is of interest both to the common users and large-scale entities such as whole governments and businesses. One of the most important aspects of energy is crude oil, which is the most consumed energy commodity in the world, as indicated by Alvarez-Ramirez et al. (2003).

In some studies, machine learning models such as the support vector machine have been compared in performance to earlier statistical methods such as the ARIMA model as applied by Xie et al. (2006) to predict crude oil prices. In this particular study, the best performance of the support vector machine algorithm was attributed to its nature of being immune to overfitting, which is a common problem in machine learning algorithms. This is possible because the support vector machine model is trained as a convex optimization problem that efficiently models nonlinear associations, as described by Xie et al. (2006).

When forecasting the wind energy resource, consideration must be given to the fact that it is a highly changing resource, as described by Pinson et al. (2007). Studies using statistical probabilistic strategies to model both the market demand and sentiments around wind power generation constraints have been conducted. For example, in a study conducted by Pinson et al. (2007), both bidding approaches and the electricity pool factors were used together with fuzzy sets in a probabilistic forecast. While some probabilistic forecasts are constructed based on Gaussian processes that utilize the Bayesian model, others have been implemented using vector autoregressive models as indicated by Cavalcante et al. (2016).

Advanced approaches, such as the ensemble deep learning method proposed by H. Z. Wang et al. (2017), have shown significant accuracy while minimizing the effects of data noise and

inconsistencies that exist in wind power. Other deep learning methods consist of the long short-term memory neural network and the Elman neural network as applied by Liu et al. (2018) to forecast the wind speed. This approach leveraged the long short term memory model's ability to significantly predict low-frequency data while the Elman neural network worked best on high-frequency data as applied by Liu et al. (2018). The Gated Recurrent Unit and Convolutional Neural Networks are other commonly applied deep learning methods which have displayed great performance in predicting energy dynamics as described by Alkhayat and Mehmood (2021).

Environment

Aspects to deal with environmental forecasting consists of factors such as prediction of carbon dioxide emissions, weather, the quality of air, floods and other related natural environmental calamities. Extensive research has been funded in this area owing to the fact that there has been a prevailing climatic threat in the world today since the industrial revolution. Earlier approaches used in forecasting carbon dioxide include the econometrics method, which involves spatially related variables as applied by Zhao and Du (2015).

In other studies, both the statistical and deep learning methods are integrated into a hybrid strategy. For example, Zhao (2017) applied both the regression model and a back propagation neural network to predict the carbon dioxide emissions in the United States of America. Other statistical approaches such as the Gaussian processes regression have also been proposed and shown to improve results of the classical back propagation neural network approaches as applied by Fang et al. (2018) to predict the carbon dioxide emission.

Weather forecasts are conducted daily by meteorologists in multiple countries. The most commonly applied approach includes the Numerical Weather Prediction approach alongside other techniques in a hybrid approach, as applied by Tiwari et al. (2018) to predict solar irradiance alongside the Gradient Boost Regression algorithm. Other machine learning algorithms have also been proposed for predicting weather factors. Algorithms such as the K-means clustering algorithm have been proposed alongside decision tree classifiers as applied in the study conducted by Murugan Bhagavathi et al. (2021) to predict the weather using a feature vector that consolidates weather elements including the temperature, pressure, humidity, wind speed, rainfall, snowfall, and snow depth.

As Das et al. (2022) explain, when forecasting certain environmental elements such as flooding, a multi-perspective is needed to consider advanced model calibration strategies such as the use of remote sensors in hydrodynamic models that take into account several feature variables such as water level and river flow. In addition to advanced system support and integration, real-time data sources such as wind flow patterns, water quantity in the soil, and evaporation need to be considered when feeding artificial intelligence models such as artificial neural networks as described by Piadeh et al. (2022). In other studies, hybrid approaches that incorporate both recurrent neural networks and the ARIMA model in a real-time perspective have been proposed to forecast flooding (Cai & Yu, 2022).

Health

In the health sector, forecasting techniques have been proposed and applied in a variety of areas that include healthcare impact, supply chain management of clinical resources, mortality, fertility, pandemics and epidemics, and so many others as described by Petropoulos et al. (2022).

When modeling epidemics for prediction contexts, approaches such as contact network epidemiology are applied as described by Meyers et al. (2006). According to this approach, the nature of contact from one individual to another in a given societal setting is extracted. Using such parameters, semi-directed contact networks are constructed to predict the epidemic as applied by Meyers et al. (2006). When constructing these epidemiological network models, it is important to note that making a more complex model does not imply better results, as described in the research by Blower (2011).

Most directed contact network models have the limitation of failing to consider the heterogeneity factor that exists in the contact weights. In order to eliminate such limitations, Kamp, (2012) proposed a weighted network model that puts into perspective the epidemiological elements, which include the rate of early epidemic expansion and the basic reproductive ratio. The nature of epidemics can be likened to several network structures. Liang et al. (2013) considered the scale-free network strategy, which intertwined new variables in the modelled problem. These variables include time-shifting functions and probabilistic elements that model the concentration and rate of infection of the epidemic, respectively.

According to research conducted by Nsoesie et al. (2013), regarding the prediction of the influenza epidemic, methods such as time series models and compartmental models have been widely applied. The ARIMA model is among the time series models used to forecast the epidemic. On the other hand, compartmental models include the susceptible–infectious–recovered (SIR) and susceptible–exposed–infectious–recovered (SEIR) models. According to Nsoesie et al. (2013), in these models, the sample population is segmented depending on the infection states. The rates at which people move from one segment to another are then obtained and modeled. It is also worth noting that the major setback with the compartmental models is that they cannot extract certain contact patterns in some situations.

Yang et al. (2014) sought to improve the SIR model when predicting the influenza epidemic by integrating it with several filters to better estimate the model's parameters. It was observed in this study that the ensemble filters and the basic particle filters provided the best results. Caudron et al. (2015) proposed the time series SIR model (TSIR) that is based on a stochastic nature and uses difference equations to predict measles. This model was observed to be more appropriate than other SIR models when dealing with smaller populations and also in circumstances whereby the measles infection shows little seasonality.

Wu et al. (2015) conducted a study in which two hybrid algorithms were deployed to predict the hemorrhagic fever with renal syndrome. In this approach, the models used include the nonlinear autoregressive neural network, the generalized regression neural network, and the ARIMA models. The ARIMA model was used as a baseline strategy to compare the performance of the other two models. During the neural network training phase, the Levenberg-Marquardt algorithm was used. A sequential approach was followed whereby the predicted outputs by the ARIMA model with their corresponding time factors were used as inputs for the generalized regression neural network.

J. Wang (2017) proposed a satellite-based prediction strategy for the oyster norovirus outbreak. In this method, the main components designed include the satellite models, an artificial neural network called NORF, and lastly, the mapping criteria. The satellite models were used to generate data regarding the environmental state by considering factors such as temperature. The NORF algorithm was used to forecast the potential for an outbreak of the norovirus to occur. The mapping methods were used to display graphically the spatial concentrations of the outbreaks in various areas.

In order to circumvent the challenge of forecasting epidemic outbreaks in real time, Funk et al. (2018) proposed a semi-mechanistic model that predicts the West African Ebola outbreak in real time. The proposed approach was based on the widely applied SEIR model with a modified version. To obtain parameter samples and model states, the Metropolis-Hastings particle Markov chain Monte-Carlo system was used. With this system, a total of 10,000 sample features were obtained with an assumption based on the "no change" premise. According to the results of this study, it was concluded that the mechanistic models perform so well on short-term forecasts in contrast to long-term forecasts. Also, the underlying stochastic nature of the Ebola epidemic spread enabled the proposed model to match quite well with the epidemic trajectories that were being studied.

Chowell et al. (2019) applied a sub-epidemic modeling architecture that was based on the generalized-logistic growth model in order to improve existing growth models that can only deal with forecasting epidemics with only one peak. Three epidemics consisting of severe acute respiratory syndrome (SARS), Ebola, and plague were studied in three countries. The proposed was used to capture single-peak, endemic, and epidemic waves that displayed fluctuating behavior. During parameter estimation, the parametric bootstrap approach was used to calculate the parametric uncertainty factor.

Adhikari et al. (2019) proposed a deep learning recurrent neural network model called EpiDeep to forecast epidemic outbreaks, especially influenza related epidemics. Both query length and full-length clustering techniques were used alongside the input data encoder and decoder components. The embedding mapper element was used to provide the final embedded data from the query length data clustering components. The embedded data was used to interpret the epidemic’s intersections in terms of the seasons it will occur with respect to previous occurrences. Encoder and decoder elements were used to obtain the present season’s existence and also to integrate outputs from clustering phases for forecasting respectively.

To perform short-term COVID-19 forecasts in 10 Brazilian states, Ribeiro et al. (2020) used heterogeneous machine learning techniques such as ridge regression, stacking-ensemble learning, autoregressive integrated moving average, cubist regression, random forest, and support vector regression algorithms. As a meta-learner, the Gaussian process was used, while foundation learners included random forest, ridge regression, and other techniques. A

multi-day future forecasting technique was used to guarantee that the model's performance can be tracked in diverse circumstances.

P. Wang et al. (2020) applied both the logistic and Prophet models to forecast the COVID-19 epidemic trend in countries like Brazil, Russia, India, Peru, and Indonesia. In their approach, the epidemiologic data was first fitted by the logistic model using the nonlinear least squares technique that updates the initialized variables. After determining the cap value of the maximum number of cases by logistic processes, it was then used by the Prophet model for fitting and forecasting with the rest of the epidemiologic data. The epidemic curves were analyzed using the epidemic peak, the fastest growth point, and the final turn points, which implied a sharp increase, a gradual increase, and a point at which recovered cases exceed confirmed cases, respectively. Results in this study showed that the proposed method was able to forecast the main turning point and epidemic size.

The SEIR and LSTM models were used by Z. Yang et al. (2020) to predict the spread of COVID-19 in China. The SIER technique was used to simulate both the COVID-19 epidemiological data and the mobility data. The operation's essential parameters β, σ and γ were identified. The product of the daily number of people who come into contact with COVID-19 patients, β, and the probability of transmission was obtained. σ was specified to be the time it takes for a COVID-19 patient to show signs and symptoms of infection. The average mortality or recovery rate was calculated as γ. The pace of pandemic propagation in Hubei province was calculated using these criteria. The LSTM model was then fed these parameters as input.

Shahid et al. (2020) applied two statistical models and three deep learning models to predict the COVID-19 cases for ten countries: Brazil, Germany, Italy, Spain, UK, China, India, Israel, Russia, and the USA. The statistical models include the ARIMA and support vector machine models, while the deep learning models are comprised of the LSTM, bidirectional long short term memory, and gated recurrent network models. Evaluation of the models was performed using three metrics. These consisted of mean absolute error, root mean square error, and r2_score indices. The best performance in this study was obtained by the bidirectional long-term memory model. The worst performance was observed with statistical models, the ARIMA and the support vector machine.

Huang et al. (2020) proposed a convolutional neural network model to perform the task of forecasting the COVID-19 confirmed cases in China. For comparison purposes, the LSTM,

Gate Recurrent Unit, and Multilayer Perceptron models were used. For the proposed model, the input data was processed into a three-dimensional matrix while the output data was a two-dimensional matrix. At the training phase, both the weight and repeat operator parameters were initialized as 0. Batches were then used for training. From each batch training operation, the weights and error values were analyzed and compared. According to the obtained results in this study, the convolutional neural network outperformed the rest of the models.

Zeroual et al. (2020) proposed a comparative approach that is comprised of five deep learning prediction models to predict new and recovered COVID-19 cases in six countries, including Italy, Spain, France, China, the USA, and Australia. The proposed models included the simple recurrent neural network, long short-term memory, bidirectional LSTM, gated recurrent units, and variational autoencoder models. Using performance metrics such as the RMSE, MAE, and MAPE, the best results were obtained with the variational autoencoder model. The variational autoencoder model was observed to be able to extract the largest percentage of data variations, hence better forecasting.

Pal et al. (2020) determined COVID-19 risk categories using the LSTM model and Bayesian optimization. The search space had to be established before the hyperparameters could be obtained. The model was given the ideal hyperparameters, which it utilized in the local trend prediction phase to provide country-specific forecasts. Finally, a fuzzy rule-based risk classification procedure was utilized to identify each country's risk level, using data gathered from the preceding module. This study found that weather had no major influence on the propagation of COVID-19.

COVID-19 prediction and comparison analysis were studied by Shastri et al. (2020), utilizing versions of long term short-term memory neural network models. Bidirectional long short-term memory, convolutional long short-term memory, and layered long short-term memory were among the models studied. Two countries were chosen as case studies. The United States and India are two of them. Tools like MinMaxScaler were employed to do data normalization since models are sensitive to the size of data input values. The magnitude of the COVID-19 problem was used to classify different areas of the United States and India into categories. The basic, intermediate, and serious categories were used in this perspective. Severe regions were defined as those with a large number of COVID-19 cases.

The ARIMA statistical model was used by Gebretensae and Asmelash (2021) to predict the spread of COVID-19 in Ethiopia. The autocorrelation function (ACF) and partial autocorrelation functions (PACF) were utilized to derive the most efficient and effective models. This study included model diagnosis, stationarity testing, and parameter testing. Gebretensae and Asmelash (2021) also employed the Box and Jenkins least squares method to approximate the model's parameters accurately. The Root Mean Square Error (RMSE) statistic was used to examine the prediction models. ARIMA (0, 1, 5) and ARIMA (0, 1, 5) were the most effective models (2, 1, 3).

To predict the positive COVID-19 result in a PCR test, Zoabi et al. (2021) used the gradient-boosting method in combination with the Sharpley Additive explanations (SHAP) bee swarm plot. Sex, contact with COVID-19 patients, and the existence of the five most noteworthy COVID-19 symptoms were among the basic variables recovered for the model's training phase in this technique. Decision-tree learners were incorporated into the primary model. The information used in this study comprised all the previous data from Israel's nationwide COVID-19 tests. For more accurate results, deep learning tactics such as early termination of the algorithm were applied using the validation dataset. The model's ability to differentiate between positive and negative situations was measured using the area under the receiver operating characteristics metric (AUROC).

Marzouk et al. (2021) applied the LSTM model, convolutional neural network, and multilayer perceptron neural network models to predict the COVID-19 epidemic in Egypt. Model fitting and assessment were done on data obtained from 4 February 2020 to 15 August 2020. The data was preprocessed by being given to the models. Sub-sequences of 20 days were used while data scaling of a range of 0–1 was used. For all, the maximum number of epochs was 200 and the minimum batch size was 256. The best performing model with the smallest root mean square value was the LSTM model. The second phase involved using the LSTM model to forecast cases for a month ahead of time.

Hssayeni et al. (2021) integrated the daily mobility data with the COVID-19 case value to predict the cumulative number of cases using the LSTM model. This model was selected in order to determine the impact of how the historical, present number of cases, and mobility data affect future COVID-19 cases. A grid search criterion was employed using the validation dataset in order to determine the optimal number of hidden layers for the LSTM model. It was observed in this study that most of the studied counties had a correlation

greater than 0.7. Also, it was observed that the case number is inversely proportional to the retiree age demographic and a direct proportionality with the young age demographic.

Yu et al. (2021) proposed an online artificial intelligence system to examine the spread of the COVID-19 pandemic. In this approach, two datasets were combined. These included the Oxford and John Hopkins university datasets. Both deep learning and statistical models were developed to perform the prediction task. Deep learning methods include the feedforward neural network, multilayer perceptron neural network, and the LSTM models. On the other hand, statistical models included the ARIMA model. The best performance was provided by the LSMT model. The proposed approach provided daily dynamics of the COVID-19, which included visualization and predictions for 171 countries.

2. PREDICTION METHODS

In this section, the most commonly applied prediction methods, technologies, and systems are explained in detail. Prediction is founded on the assumption that present and previous information may be used to create future projections, as described in research by Petropoulos et al. (2022). This is grounded in the fact that in past data there exist patterns and trends that, if critically analyzed, can be used to provide insights into the nature of future values. The outcomes of a prediction aren't necessarily intended to be precise future occurrences, but they might be a percentage or a range of projected values as described by Majid and Mir (2018). The major objectives of a forecast are met when the results can give significant direction on the actions that decision-makers should take. For instance, a forecast that suggests quick attention to the contextual variable of study by a particular proportion.

According to Majid and Mir (2018), in general, there are two techniques for forecasting: qualitative and quantitative. To develop predictions, qualitative research approaches rely on expert opinion from specialists in certain disciplines. Quantitative approaches, on the other hand, infer recent and historical activity into the future using past information and a predictor. Quantitative techniques are divided into two categories: time-series approaches and econometric approaches. According to Petropoulos et al. (2022), a forecasting problem can be univariate or multivariate depending on the variables employed. A single feature variable is utilized to describe the problem for simulation or learning methods in a univariate forecasting challenge. This is predicated on the assumption that the selected feature variable can adequately represent the variable under investigation.

In most cases, however, forecasting problems require the modeling of many feature variables. This is due to the fact that the variable under investigation is better represented or that its distribution is influenced by a number of factors that all play a part in the modeling process. The basic steps commonly followed during forecasting comprise of five main stages. These include, problem identification, accumulating data, preliminary analysis, model selection and fitting, model usage, and assessment as described by Hyndman and Athanasopoulos (2022).

In the information accumulating or gathering phase, either statistical or data from experts' knowledge is obtained. This data is then analyzed for any structures such as trends, cycles, and seasonality. Given the analyzed data, the best prediction method, technology, or model

is selected, fitted, assessed, and finally used for forecasting. Prediction methods can be broadly grouped into statistical and deep learning approaches.

2.1. Statistical prediction methods

Statistical approaches use major statistical computations to derive patterns in any given data to make forecasts. These methods broadly include commonly used approaches such as linear regression, nonlinear regression, exponential smoothing, ARIMA and many others.

2.1.1. Simple Linear regression prediction

Regression is a statistical approach for modeling and studying the connections between one or more forecast or classifier variables and a result or target value. A regression analysis study's end result is frequently the creation of a model that can be used to predict or forecast potential values of the dependent variables based on the values of the predictors as described by Montgomery et al. (2015).

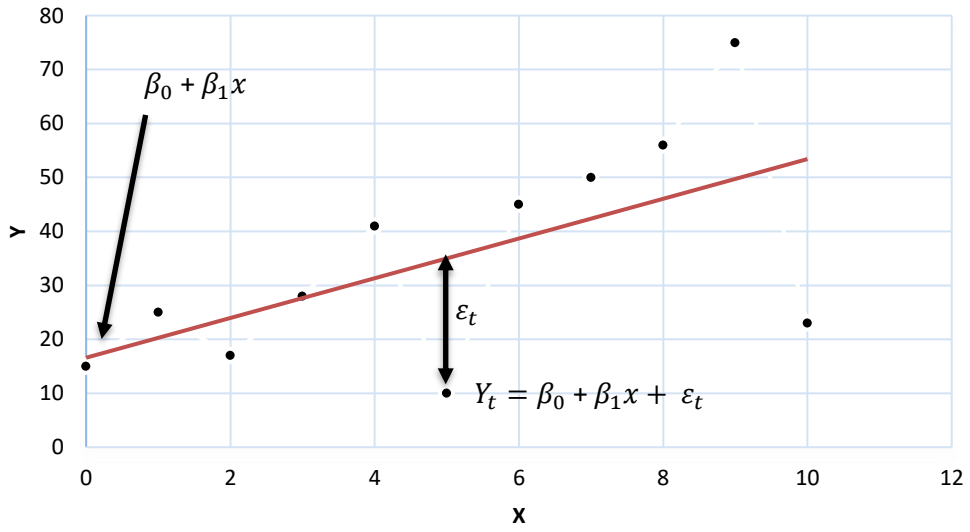

Figure 2.1. Sample representation of the simple linear regression model.

As displayed in figure 2.1 above, a simple linear regression model is used to extract the linear relationship between the forecast variable and the predictor variable. The forecast and predictor variables have been represented by Y and x, respectively.

$$Y_t = \beta_0 + \beta_1 x + \varepsilon_t \quad (2.1)$$

In both figure 2.1 and equation 2.1, β_0 and β_1 represent the intercept and slope of the simple linear model curve respectively as explained by Hyndman and Athanasopoulos (2022). On the other hand, ε_t represents the error terms of the linear model. This error factor represents the amount of deviation of the predicted values from the actual values, as depicted by the straight line in the graph.

According to Hanck et al. (2021), the reasons for the existence of these deviations range from the absolute randomness of the dependent variable to the fact that it might be due to the omission of additional elements that are important in understanding the dependent variable or experimental errors. The linear regression forecasting approach has shown significant performance in the prediction of trends in the given data distribution, as applied by Lin and Tsai (2015) to predict the mortality rates across the age demographic. This is a robust technique that is much needed when making long term decisions and planning.

2.1.2. Multiple linear regression prediction

When performing forecasting of the dependent variable using linear regression models, in some cases, there is a need to incorporate several predictor variables that can better describe the dependent variable. In this case, the simple linear regression method described initially cannot be used since it considers only a single predictor variable. This brings forth the need and usage of multiple linear regression approaches for better results in these cases. Multiple linear regression models take into perspective the modeling of a variety of predictor variables as described by T. Fang and Lahdelma (2016).

$$Y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_k x_{k,t} + \varepsilon_t \quad (2.2)$$

In equation 2.2 above, the multiple linear regression model form has been mathematically depicted as described in a study by Hyndman and Athanasopoulos (2022). In this equation, the forecast variable is represented by the term Y_t which represents a particular dependent variable that is forecasted by the model at a particular time t. β_1,..., β_k represents the coefficients that depict the combined effect of all other predictors on the outcomes of the dependent variable values. On the other hand, the terms x_1,.. x_k represents the k predictor values.

To ensure that the predictions provided are accurate, the correlation between the errors and the independent variables' observations should be zero, with the latter exhibiting minor

autocorrelation as described by Petropoulos et al. (2022). Other assumptions for better results of a linear regression model prediction comprise the need for a normal distribution of the residuals while having a constant value for the extent of how far experimental values vary from the average of predicted values and a need for the residual value to have a mean zero as described by Hyndman and Athanasopoulos (2022).

In cases where there exists no significant linear relationship between the dependent and independent variables, linear regression models have shown poor prediction results, as observed in a study by Choubin et al. (2016).

Estimation of coefficients in linear regression prediction models

In this section, the strategies used to obtain the unknown parameters in linear regression prediction models are discussed in detail. These approaches include widely applied methods such as ordinary least squares, weighted least squares, and maximum likelihood estimation.

According to Weisberg (2013), the ordinary least squares estimation method involves the selection of model coefficients in a manner that reduces the residual sum of squares. When dealing with simple linear regression models, the calculations required for least squares rely just on the means of the parameters and their summation of squares and summation of cross-products. This is normally achieved by removing the average from each of the numbers before squaring or doing cross-products (Weisberg, 2013).

$$\sum_{t=1}^{T} \varepsilon_t^2 = \sum_{t=1}^{T} (y_t - \beta_0 - \beta_1 x_{1,t} - \beta_2 x_{2,t} - \cdots - \beta_k x_{k,t})^2 \quad (2.3)$$

According to Hyndman and Athanasopoulos (2022), in a process called fitting, by reducing the summation of the squared errors, the least squares concept allows for efficient parameter selection as depicted in equation 2.3. This implies that the terms $\beta_0, \beta_1, \ldots, \beta_k$ are selected so as to reduce the squared error term ε_t^2 as mathematically illustrated in equation 2.3.

On the other hand, the weighted least squares estimation method is an extension of the ordinary least squares that incorporates information about the variability of the observed into the regression as described by Wikipedia contributors (2022). According to Kantar (2015) and Glen (2021), the weighted least squares estimation method needs to be applied when the variances in a given observation are not similar, which describes a situation whereby the data observations do not comply with the concept of homoscedasticity.

$$L = \sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i)^2 \quad (2.4)$$

According to Montgomery et al. (2015), the weighted least squares estimate can be mathematically obtained from equation 2.4 above. In this equation, w_i represents the weight, y_i and x_i represents the independent and dependent variable matrices, respectively. The unknown model parameters are represented by terms β_0 and β_1.

The weighted least squares estimate is ideal for getting the most out of tiny data sets and is the only way to deal with data points of different quality, as explained by Glen (2021). When using the maximum likelihood estimate, the objective is to estimate the model's unknown coefficients in such a way that the probability of obtaining the actual sample values is made maximum while the likelihood function, which is the joint probability distribution of the observations considered as a function of the unknown coefficients, is used to calculate this probability (Hanck et al., 2021). This implies that values that enable the model to make the most likely predictions when compared to the actual values are the most optimal model coefficients.

Hybrid linear regression model to forecast heart illness

Heart issues are among the most frequent ailments, and doctors must diagnose them as early as possible in order to treat them and increase the chances of the victims' survival. Cardiovascular diseases are one of the greatest reasons for death all over the world, as described by Srinivas et al. (2020). The health industry has stored information that aids in making early assessments by applying various forecasting approaches to detect underlying illnesses such as coronary artery disease.

Srinivas et al. (2020) suggested a Hybrid Linear Regression Model with two phases of implementation. Missing values are imputed using the K-nearest neighbor method and simple mean imputation during data preparation. The highest contributing factors to the cause of the illness are then extracted using the Principal Component Analysis strategy. Furthermore, the inherent correlation between the predictor and forecast variables is extracted using the stochastic gradient descent technique (Srinivas et al., 2020).

Supported by the empirical association with the prediction of heart disease, a collection of illness factors $A_k, k = 1,2,\ldots n$ is established, whereby n represents the number of these kind of factors (Srinivas et al., 2020). The data consisted of the $A_k(x_i)$ parameter inputs,

where $W_k, k = 1,2,\ldots,n$ is a collection of weight vectors for every one of the N illness factors used in the prediction process, while the number of individuals with the disease is modelled by x_i. The algorithm's goal is to estimate the likelihood that a given patient would be diagnosed with heart disease, as well as the weighted values of the variables used to compute the likelihood (Srinivas et al., 2020).

During parameter tuning, four model coefficients are employed in the hybrid linear regression model. These coefficients consist of the iteration number, loss and penalty functions, and the rate of learning (Srinivas et al., 2020).

$$S = \sum_{i=1}^{n} |y_i - (f(x_i))| \quad (2.5)$$

In equation 2.5 above, the absolute deviation S between the target data points y_i and the expected data units $f(x_i)$, is mathematically denoted. The penalty function is applied to this metric in order to reduce the number of deviations (Srinivas et al., 2020).

Comparative prediction of healthcare costs using regression algorithms

In a study conducted by Thongpeth et al. (2021), four notable regression algorithms were proposed to forecast the costs incurred by patients in hospital visits owing to chronic illnesses. These models are comprised of Least Absolute Shrinkage and Selection Operator (LASSO) regression, Ridge regression, Elastic net regression, and Support Vector Regression.

LASSO regression is a commonly applied modelling algorithm in the health sector. LASSO makes use of contraction to achieve linear regression. Compressibility happens whenever a data point decreases towards a focal point, such as the mean. As an outcome, it works well with algorithms that have a lot of correlation (Chintalapudi et al., 2022). It also automates algorithm selection processes like parameter exclusion and feature extraction. In clinical applications, the LASSO regression model is increasingly being utilized to forecast infection outcomes and adverse reactions (Chintalapudi et al., 2022).

By decreasing the scores of less informative factors closer to zero, Lasso integrates a normalization strategy to minimize the overfitting problem and the algorithm's complex nature, as described by Thongpeth et al. (2021). The technique has been used in the areas of brain simulation, biomarker screening, healthcare expenditure projection, and myocardial infarction early diagnosis (Chintalapudi et al., 2022).

$$\beta' = \underset{\beta}{argmin}\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|\}, \lambda \geq 0 \quad (2.6)$$

In equation 2.6 above, the LASSO coefficient β_j can be mathematically computed by optimizing the objective function β'. The shrinkage factor λ denotes the quantity of shrinkage needed.

The Ridge regression element in this application is employed to ensure that the total complexity of the algorithms is minimized. This nature of minimization is also incorporated with the L2 normalization techniques to ensure better results as described by Thongpeth et al. (2021). This strategy makes use of the summation of squared weights that is directly used to model the penalty coefficient.

$$\beta' = \underset{\beta}{argmin}\left\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}, \lambda \geq 0 \quad (2.7)$$

In equation 2.7 above, the Ridge coefficient β_j can be formally calculated by adjusting the objective function β'. The shrinkage element λ signifies the magnitude of shrinkage required.

The elastic net regression is derived by taking an integral unit of both the LASSO and ridge regression components. With this nature of combination, the resultant bias factor obtained is significantly smaller than the values obtained from both LASSO and ridge regression units as described by Thongpeth et al. (2021). On the other hand, the support vector regression element makes use of nonlinear alterations in complex data contexts using an assortment of hyperplanes (Thongpeth et al. 2021). This allows this model to provide optimal results by making sure that the hyperplane gap between the forecasted values and observed data points is significantly lower.

$$f(x) = w.x + b \quad (2.8)$$

The operation of the support vector regression model can be mathematically derived as represented in equation 2.8 (Thongpeth et al. 2021). According to this equation, x denoted to input set of values while w represents the set of weights applied to the input vector values. On the other hand, the gap between predicted and original values is denoted by the error term b.

2.1.3. Nonlinear regression prediction

Nonlinear regression prediction methods involve observational data that is represented by a function that relies solely on a single or even more predictor factors and is a nonlinear composition of regression coefficients. This form of regression enables the generation of models with any connections between predictor and dependent variables, contrasting with typical linear regression, which would be limited to evaluating only linear regression models as described in a study conducted by Yasar et al. (2012).

Although linear regression models are effective in certain situations, nonlinear functional forms are more appropriate in others. Transforming the forecast variable and/or the predictor variable using approaches such as logarithmic transformation before evaluating a regression model is the easiest technique to model a nonlinear connection, according to Hanck et al. (2021) and Hyndman and Athanasopoulos (2022). Although this results in a non-linear functional form, the model's coefficients remain linear as described by Hyndman and Athanasopoulos (2022).

Logarithmic transformation

During logarithmic transformation, the fluctuations extracted from the variable values are converted into percentage fluctuations using logarithms. These logarithms take into consideration three formats as described in the study conducted by Hanck et al. (2021).

$$Y_i = \beta_0 + \beta_1 \times In(X_i) + u_i,\ i = 1, \dots, n. \tag{2.9}$$

The first case deals with the logarithmic transformation of the independent variable X without transforming the dependent variable Y. In equation 2.9, this kind of logarithmic transformation of the independent variable X is mathematically illustrated. The u represents the error term. In this perspective, when there is one percent fluctuation in the independent variable, the expected change in the dependent values is expected to be 0.01 x β_1 as described by Hanck et al. (2021).

$$In(Y_i) = \beta_0 + \beta_1 \times (X_i) + u_i,\ i = 1, \dots, n. \tag{2.10}$$

On the other hand, the second scenario is the opposite of the first scenario whereby the logarithmic transformation is performed on the dependent variable $Y$and not on the independent variable X as illustrated in equation 2.10. When a one-unit change occurs in the

independent variable's values, the change in Y is expected to be $100\beta_1$ percent (Hanck et al., 2021).

$$In(Y_i) = \beta_0 + \beta_1 \times In(X_i) + u_i,\ i = 1, \dots, n. \quad (2.11)$$

Lastly, the third case involves the logarithmic transformation of both the dependent and independent variables. This kind of transformation is mathematically depicted in equation 2.11. After obtaining the logarithms of both dependent and independent variables, when the values of X fluctuate by one percent, the values of the dependent variable are expected to change by β_1 percent (Hanck et al., 2021). Nonlinear models have several merits, including internal consistency, understandability, and forecasting. Nonlinear models may accommodate a wide range of average functions in theory, while every nonlinear model is much less versatile in terms of the types of data it can explain than linear models, and also forecasts are more reliable than comparable polynomials, notably when data is not available as described by Archontoulis and Miguez (2015).

Predicting epidemics with method of Analogues

In the framework of variational systems theory, the methodology of analogues has first been introduced as a nonlinear predictive tool (Moore & Little, 2014). The method uses delay co-ordinates to encapsulate the time series in a feature space with local estimation to derive the non-linear function. It is examined for examples to utilize as predictive variables instead of using the training dataset to compute the algorithm's coefficients. This model utilizes previous series that are related to the normal series and makes a forecast using their predecessor values (Moore & Little, 2014). Viboud (2003) proposed the application of the method of analogues to forecast the spread of the influenza epidemic. In this application, both univariate and multivariate time series perspectives are considered in the modeling phase. The simplest of these phases deals with the forecasting of the national occurrences of influenza-like infections.

$$dist\big(X(T), X(t)\big) = \sum_{j=0}^{l} (I(T-j) - I(t-j))^2, t < T \quad (2.12)$$

This method makes use of the distance magnitude between the occurrences at a particular time with the past observations to extract the best matching data points within a given range of time (Moore & Little, 2014). This can be mathematically observed in equation 2.12 above (Viboud, 2003). In this specific application of predicting influenza, in equation 2.12, I

denoted the observed influenza incidences at a given time t in weeks. T indicates the current week beyond which the forecast is to be made (Viboud, 2003). Weighted mean metrics are then computed to determine the closest neighboring data points for consideration in the future forecast vectors (Moore & Little, 2014).

$$dist\big(X(T), X(t)\big) = \sum_{k=1}^{21} \sum_{j=0}^{l} (I_k(T-j) - I_k(t-j))^2 \quad (2.13)$$

In equation 2.13, the multivariate regional model has been mathematically represented (Viboud, 2003). It is evident that the second phase of this application extends the initial forecasting model used to project future values on a national level. In this equation, the regional element is modelled using the k coefficient. In this phase, the weighted mean vector is derived to compute the best matching data values in the sequence (Viboud, 2003).

2.1.4. Exponential smoothing prediction

In broad terms, smoothing processes comprise techniques to isolate the signal and noise elements from a given data distribution, as described in a study conducted by Montgomery et al. (2015). The signal component signifies the existing patterns in the data that may be present due to the inherent nature by which this data is obtained.

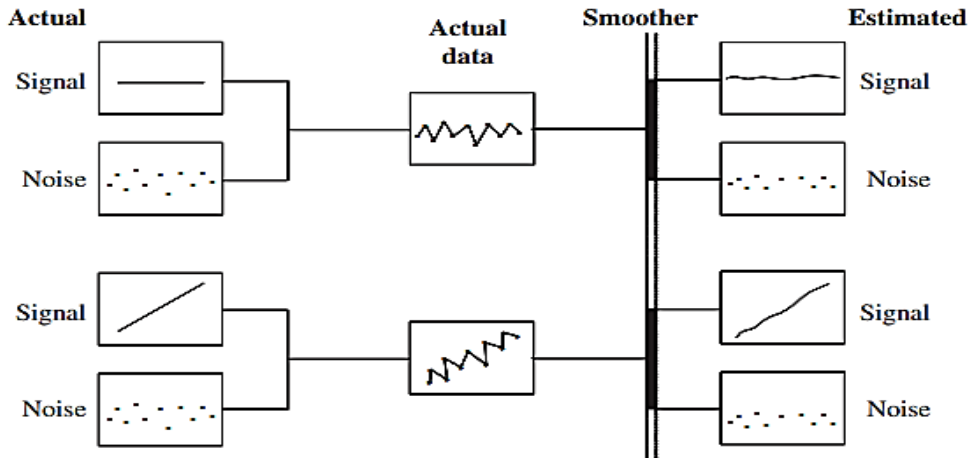

Figure 2.2. Illustration of smoothing processes.

According to Montgomery et al. (2015), smoothing tools aim to extract estimations of signals which may take different structures ranging from simple constants to comparatively more complex structures as illustrated in figure 2.2. In this figure, it is evident that smoothing techniques obtain signals from the data distributions by relating the present data points to prior values.

According to Frost (2021), exponential smoothing is a prediction approach that generates projections based on weighted averages of previous data, with the weights of earlier data values decreasing exponentially. Exponential smoothing methods achieve this by processing the weighted average of the prior and present values using reducing weights on the previous data as described by Holt (2004).

In addition, these exponential smoothing prediction techniques broaden the scope of operation to comprise data with trends and seasonal components, which allows analysts to be able to modify how soon older observations lose their significance in computations by altering model parameters, which, as a result, enables analysts to have the freedom to modify the relative relevance of new vs. old data to meet the demands of their research topic. (Frost, 2021).

Table 2.1. Classification of exponential smoothing approaches

Trend element	Seasonal element		
	N (None)	**A (Additive)**	**M (Multiplicative)**
N (None)	N, N	N, A	N, M
A (Additive)	A, N	A, A	A, M
A_d (Additive damped)	A_d, N	A_d, A	A_d, M
M (Multiplicative)	M, N	M, A	M, M
M_d (Multiplicative damped)	M_d, N	M_d, A	M_d, M

In table 2.1 above, the general classification of exponential smoothing approaches has been given where it is evident that there are only two state space algorithms; the first relates to a model that deals with the occurrence of additive errors while the second relates to a model that works on the existence of multiplicative errors, as described in a study conducted by R. Hyndman et al. (2008).

In related literature, some of the exponential smoothing approaches illustrated in table 1.1 have been largely known by different names. That is to say, methods that include, simple exponential smoothing (N, N), Holt's linear method (A, N), Additive damped trend method (A_d, N), Additive Holt-Winters' method (A, A), Multiplicative Holt-Winters' method (A,

M) and Holt-Winters' damped method (A_d, M) as illustrated by R. J. Hyndman and Athanasopoulos (2022), Majid and Mir (2018) and R. Hyndman et al. (2008).

Simple exponential smoothing (N, N)

When data being processed for a prediction task lacks both the trend and seasonal elements, this exponential smoothing approach is the ideal option. For a given time series, $y_1, y_2, \ldots, y_n$, the forecast for the next time period t can be obtained as represented mathematically in equation 2.9 below using the simple exponential smoothing approach of R. Hyndman et al. (2008).

$$\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t) \tag{2.14}$$

In equation 2.14, the assumption is considered of having to predict the next value y_t when the presently obtained data is bounded within $t - 1$ time inclusively. The α factor is a constant whose value ranges from zero to one according to the simple exponential smoothing criterion applied when the data displays no significant trend or seasonal patterns. When further expansion of the simple exponential smoothing function is executed, it can be evidently illustrated that the value of $\hat{y}_{t+1}$ depicts a weighted moving average after having taken into perspective the prior data values, which display an exponentially reducing weight (R. Hyndman et al., 2008).

According to Hyndman and Athanasopoulos (2022), the simple exponential smoothing procedure can be mathematically denoted in a totally different way which captures both the smoothing and forecast mathematical representations as illustrated in equations 2.15 and 2.16.

$$\hat{y}_{t+h|t} = \ell_t \tag{2.15}$$

In equation 2.15, the forecast process is represented, and it illustrates the forecasted value at any given time t+1 which lies at the approximated level at a particular t. When the value of the term h is one, the output values of the fitting process can be obtained.

$$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1} \tag{2.16}$$

Equation 2.16 represents the smoothing process that considers the application of weighted averages to the selected values of the time series. The term ℓ_t denotes the smoothed value, which is the approximated level of the time series at a given time t.

Holt's Linear exponential smoothing (A, N)

This exponential smoothing method was named after Holt, the researcher who first proposed it. It aims to extend the simple exponential smoothing method by incorporating the trend factor that is not captured by the simple exponential smoothing method. This is achieved by introducing two more factors that are used in the smoothing process, as depicted in equation 2.16, equation 2.17, and equation 2.18 below (R. Hyndman et al., 2008).

$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \quad (2.16)$$

In equation 2.16, the level element of a given time series can be mathematically derived using the gradient and first smoothing parameters. This equation ℓ_t represents the time series approximation at any given time t. The gradient of the time series at a particular time t can be denoted by b_t in equation 2.17.

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\, b_{t-1} \quad (2.17)$$

In equation 2.17, the expansion of the gradient term has been represented. In this equation, the second smoothing parameter, β^*, is introduced. The b_t gradient term is obtained by computing weighted average values of prior slope values b_{t-1} which together give the final approximation of the gradient value from the variance that exists between consecutive levels.

$$\hat{y}_{t+h|t} = \ell_t + b_t h. \quad (2.18)$$

Equation 2.18 denotes the forecast component of the Holt's linear method of exponential smoothing that combines both the growth and level elements.

Damped trend exponential smoothing (A_d, A)

The damped trend exponential smoothing method was proposed to streamline the performance of Holt's linear exponential smoothing method, which had weaknesses such as a constant trend that continuously continued the forecast indeterminately. The damp trend exponential smoothing approach attains this improvement by a new factor that performs the dampening control of the trend to a uniform flat track, as described by Hyndman and

Athanasopoulos (2022). This process can be mathematically denoted as represented in equation 2.19, equation 2.20, and equation 2.21 (R. Hyndman et al., 2008).

$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \quad (2.19)$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1} \quad (2.20)$$

$$\hat{y}_{t+h|t} = \ell_t + (\phi + \phi^2 + \cdots + \phi^h) b_t \quad (2.21)$$

In equations 2.19 - 2.21, it is clear that the damped trend exponential smoothing prediction method incorporates a new dampening factor ϕ, which is applied to the forecasted value y_{t+1} at a given time period of t. The growth factor for this particular forecast is then denoted by ϕb_t. The dampening factor lies between zero and one, just like the rest of the smoothing parameters (R. Hyndman et al., 2008).

The output of this exponential smoothing method displays stationarity in the first level of differentiation instead of the second level as applied by the Holt's linear method as described by McKenzie & Gardner, (2010). This implies that when the time series data has a consistent and large existing trend, the dampening factor would need to be set to values close to one. On the other hand, for noisy data it would need to be close to zero, which results in a dampened forecast.

In cases where the data displays a significant amount of seasonality, the Holt-Winters' method was proposed to handle the seasonal factors. In this approach, three distinct factors are considered. These factors include the level, trend, and seasonality factors used in the mathematical equations. Holt-Winters' methods are broadly grouped into two groups, which are based on the nature of how the seasonality factor is considered, either in an additive or multiplicative format (R. Hyndman et al., 2008).

Exponential smoothing techniques have been extensively applied in the health sector. Some of these applications include forecasting the COVD-19 fatalities (Oladunni et al., 2021), prediction of the mosquito-borne viral illness called dengue applied by Attanayake et al. (2020), forecasting of mortality rates by Adeyinka and Muhajarine (2020), and prediction of brucellosis cases by Guan et al. (2018).

Hybrid approaches have been proposed that combine the predicting power of exponential smoothing methods with other statistical forecasting methods. These methods include

applications such as the prediction of COVID-19 transmission dynamics using the hybrid grey exponential smoothing model by Seneviratna and Rathnayaka (2022). This hybrid model combines both the double and grey exponential smoothing modelling approaches to forecast the pandemic in a phased approach.

2.1.5. Autoregressive Moving Average prediction

The ARMA model is based on the idea that the specifics in previous historical time series data can be used to forecast potential outcomes. This algorithm produces a high-accuracy outcome in a short amount of time (Meenal et al., 2022). As a linear transformation of previously observed data, this model enables the approximation of the future value of a specific time series (Gomes & Castro, 2012).

The factors of this linear transformation, which are the model's coefficients, are calculated using the time series directly, so that every data unit in the time series is best expressed by a linear transformation of some of its past values. This operation relates to the algorithm's training stage, whereas the approximation of the next value refers to the algorithm's prediction stage (Gomes & Castro, 2012).

Because of its versatility and appropriate statistical features, the Autoregressive Moving Average (ARMA) model is frequently employed in time series modeling and prediction (Meenal et al., 2022). A simple and sparse construction distinguishes the ARMA model from other commonly known statistical models, such as linear regression models (de Oliveira et al., 2022). It integrates the Autoregressive (AR) model, which regresses a parameter on its own lagged factors. On the other hand, the Moving Averages (MA) component defines the standard errors by considering the linear transformation of the parameter's own lagged factors (de Oliveira et al., 2022).

$$x_t = \emptyset_t + \emptyset_1 x_{t-1} + \cdots + \emptyset_v x_{t-v} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_w \varepsilon_{t-w}, t \in Z \qquad (2.22)$$

In equation (2.22), the formal denotation of the ARMA model has been mathematically represented (de Oliveira et al., 2022). In this equation, $\varepsilon_t \sim WN(0, \sigma^2), t \in Z$, has highlighted that the error factor is a white noise mechanism (WN) with 0 mean and variance. The ARMA(v, w) linear mathematical expression to be approximated is specified using v lags of the output and w lags of the error factor.

A least squares reduction method is commonly used to determine the coefficients of the ARMA algorithm (Gomes & Castro, 2012). By decreasing the sum of the squares of the error terms, the discrepancies seen between the values in the current data as well as the same values generated by the model can be decreased. The main concept in this methodology seeks to identify the finest model parameters from a set of values that exist in a given time series (Gomes & Castro, 2012).

2.1.6. Autoregressive Integrated Moving Average prediction

The autoregressive integrated moving average (ARIMA), as clearly explained by Ediger and Akar (2007), is a predictive analysis model widely used during time series estimation and forecasting of prospective data points in such sequences. The ARIMA model is divided into three sections: These components include the "AR," "I," and "MA."

The "AR" term refers to the autoregression component, as stated in the Corporate Finance Institute's (2022) study. This indicates that the variable in question has a linear connection between its current and previous occurrences. As mentioned in the study conducted by Investopedia (2021), an AR(1) of degree one suggests that the current data point in the series is dependent entirely on the most recent data item, but an AR(2) indicates that it is defined by two previous data units in the time series.

The integrated element, denoted by the character "I," demonstrates the number of differences between the current variables and their previous values. As detailed in the study conducted by the Corporate Finance Institute (2022), this is the element of the ARIMA model that addresses the data stationarity need for improved outcomes in ARIMA time series computation, which is achieved through differencing. The situation when the mean and variance statistical parameters in the time series data remain unchanged with regard to the time factor is referred to as stationarity in ARIMA computation.

The "MA" portion, which represents the moving average, is the final part of the fundamental ARIMA framework. As defined by Ribeiro et al. (2020), this element depicts the linear combination that occurs between the standard errors at previous periods in the time series. ARIMA is the common abbreviation for the fundamental ARIMA model (p, d, q). According to Abdulmajeed et al (2020), the p, d, and q terms reflect the autoregressive, differencing, and moving average parameters respectively.

Additional terms are added to the basic framework in a seasonal ARIMA model to account for the seasonality of the time series. ARIMA(p, d, q) (P, D, Q)m is a general depiction of a typical seasonal ARIMA model. The periodical count for each season is represented by the m element. P, D, and Q represent the autoregressive term, difference degree, and moving average for the seasonal component in the model, respectively, as described in a study performed by Noureen et al. (2019). Equation 2.17 shows the formal representation of the AR (p) factor.

$$Y_t = \delta + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t \quad 2.23)$$

In equation 2.23 above, the element Y_t signifies the value in the time series at a certain time t. The autoregression term, fixed value, and error value are indicated by p, δ and ε_t respectively.

Equation 2.24 is used to define the moving average component theoretically.

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_2 \varepsilon_{t-2} + \theta_q \varepsilon_{t-q} \quad (2.24)$$

The degree of the moving average term is represented by q in equation 2.18. Equation 2.19 could be used to find the difference factor d.

$$\Delta Y_t = Y_t - Y_{t-1} = Y_t - LY_t \quad (2.25)$$

The stationary time series value at a time interval t is denoted by ΔY_t in equation 2.25.

$$\left(1 - \varphi_1 L - \varphi_1 L^2 - \cdots - \varphi_p L^q\right)\Delta^d Y_t = \delta + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (2.26)$$

Every one of the calculations for the fundamental ARIMA model terms is combined in Equation 2.26. This was the whole ARIMA (p, d, q) model equation, with all terms derived and shown.

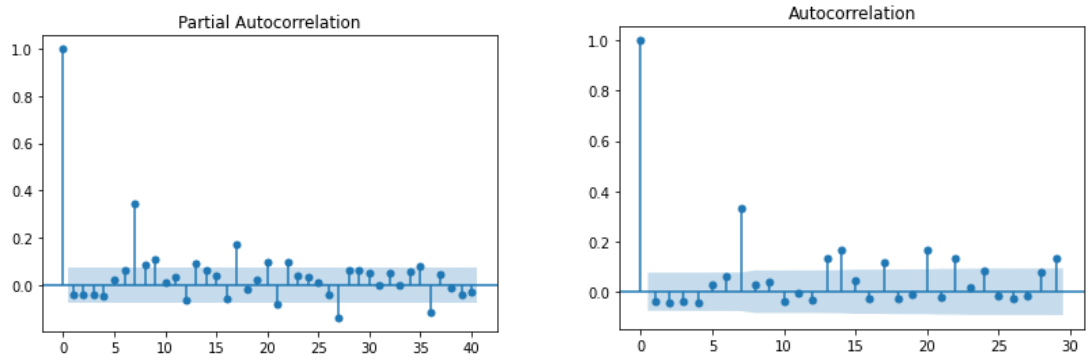

Figure 2.3. Illustration of the PACF and ACF plots

The ARIMA model's p and q factors could also be obtained using the partial autocorrelation function (PACF) and autocorrelation function (ACF) plots, as illustrated in figure 2.3. The ACF plot depicts the average correlation between data and previous values in a time series across various time lag periods. According to Noureen et al. (2019), the primary difference between the two is that the PACF identifies associations with a shorter lag time.

In statistical terminology, autocorrelation refers to the linear association that occurs between a certain variable and itself at two points in a given time frame, as defined by Mathworks (2021). The difference factor could be determined if the time series is indeed not stationary. To evaluate if the time series stationarity condition is fulfilled, the augmented Dickey–Fuller test (ADF) is employed. According to Abdulmajeed et al. (2020), the ADF analysis is used to check the null hypothesis, which indicates whether the unit root occurs in a specific time series. The Akaike information criterion (AIC) estimator has been used to find the ARIMA parameters in this study.

$$AIC = 2k - 2In(L) \tag{2.27}$$

Using the approximated model constant k and the largest number of the likelihood function L, the AIC value can be analyzed quantitatively with equation 2.27.

2.1.7. Generalized autoregressive conditional heteroscedasticity prediction

The generalised autoregressive conditional heteroscedasticity (GARCH) statistical algorithm is a general extension of the autoregressive conditional heteroscedasticity (ARCH) model, which was proposed to solve the dynamic volatility phenomeno that occurs when forecasting time series (Petropoulos et al., 2022). The GARCH model achieves this with a complex structure that incorporates the unchanging volatility factor. This allows the

possibility of the approximation of the ultimate average. The GARCH algorithm also integrates the volatility predictions that eventually take place in the final phases while adding new information (Petropoulos et al., 2022).

The GARCH algorithm includes a parsimonious depiction of the autoregression terms, which has the highest degree of computation. That is to say, a GARCH (p, q) model can be obtained mathematically as illustrated in equation 2.22 below (Mapa, 2004).

$$\sigma_t^2 = \omega + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{q} \alpha_i u_{t-1}^2 \quad (2.28)$$

Whereby, $\omega > 0,\ \alpha_i \geq 0 \quad i = 1, \dots, q \ \ j = 1, \dots, p$

It is evident in equation 2.28 above that the variance term can be obtained from a linear combination of the q lag values that resides with the prediction error factors u_t^2 which at the same time denote the ARCH coefficients, and the p lag values obtained from the prior instances of the variance terms σ_t^2 which represent the GARCH coefficients with a given constant ω that is greater than zero to avoid negative values in the computation (Mapa, 2004).

2.1.8. Epidemic prediction with deterministic compartmental models

In this section, the most commonly applied deterministic compartmental models in epidemic prediction are discussed in detail. These models analyze the epidemic's propagation while considering a nonlinear problem space as described by Alenezi et al. (2021). The main concept of these epidemic modeling and prediction models is grounded in the principle that breaks a given population into compartments (Brauer et al., 2008). This analogy considers the correlation between the rate of transmission of an epidemic from one compartment to another.

In the health sector, there are four major concepts to follow when making predictions. These include the horizon of the prediction, the assessment of uncertainties and losses, the focus and the form of data aggregation as described by Soyiri and Reidpath (2012). When dealing with uncertainties and errors, consideration is mainly given to the nature of the approach to be selected for the prediction task. This approach should be able to handle and process health data that is subject to so many uncertainties. From the validation processes, the output of the prediction procedures is expected to produce the least error possible, which is directly

proportional to the robustness of the chosen prediction technique as described by Soyiri and Reidpath (2012). The inherent nature and characteristics of a particular epidemic in transmission from one compartment to another can display significant variations because of the disparities in immunity brought about by various epidemics (Brauer et al., 2008). When constructing these algorithms as differential equations, it is considered that perhaps the infectious disease mechanism is deterministic, meaning that a population's behavior is totally defined by its past and the principles that constitute an algorithm as explained by Brauer et al., (2008).

Susceptible-Infected-Recovered (SIR) Model

Various infections, such as HIV and Ebola, have been estimated using the deterministic model. Susceptible (S), Infected (I), and Recovered (R) are three variables that SIR uses to compute the overall population (R) (Alenezi et al., 2021). Susceptible refers to the entire population that is healthier but at risk of infection. The number of people who are moderately or seriously ill is referred to as "infected." The overall number of healed and immune individuals from the outbreak, including those who died, is referred to as recovered as defined by Alenezi et al. (2021).

$$N = S + I + R \tag{2.29}$$

In equation 2.29 above, the overall population can be mathematically computed (Alenezi et al. ,2021). In this equation, N denotes the overall population, S denotes the susceptible individuals, I represent the infected individuals, and R denotes the recovered members of the population. It is clear in this equation that the overall population is just a summation of all compartments with regard to the epidemic spread.

$$\frac{dS}{dt} = -\frac{\beta IS}{N} \tag{2.30}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I \tag{2. 31}$$

$$\frac{dR}{dt} = \gamma I \tag{2.32}$$

The population is considered to be stable in the SIR mathematical formulation, with hardly any deaths or births throughout the disease forecast timeframe. The variations in S, I, and R are calculated utilizing differential equations as depicted in equations 2.30 to 2.32 above (Alenezi et al.,2021). In these equations, β denotes the daily rate of infection by the

susceptible population while γ depicts the rate of recovery, which considers the immunity factor in the recovery process.

$$R_0 = \frac{\beta}{\gamma} \tag{2.33}$$

In 2.33 above, the reproduction factor of a given epidemic is mathematically denoted (Alenezi et al., 2021). This metric uses both the rate of infection and rate of recovery terms.

Susceptible-Exposed-Infected-Recovered (SEIR) Model

The SEIR model is among the most basic epidemiologic compartmental models and is a modified version and extension of the SIR model (Alenezi et al., 2021) (Efimov & Ushirobira, 2021). It is often a widely accepted model in a variety of contexts. The SEIR model depicts the evolution of the relative proportions of four types of people in a population of fixed dimension as described by Efimov and Ushirobira (2021). These include the susceptible people S, who are capable of getting an infection while being contagious, the exposed E, and the symptomatic I contagious, who can pass the illness further to susceptible people. On the other hand, the last compartment is the recovered R, who is completely invulnerable after recovery or death. If the mortality rates are particularly important, a section D is often added (Efimov & Ushirobira, 2021).

$$N = S + E + I + R \tag{2.34}$$

The mathematical computation of the overall population in a SEIR model does not differ much from that of the SIR model as depicted in equation 2.34 above (Pandey et al., 2020). In this equation, it can be evidently observed that the added factor in this context is the exposed compartment, which is denoted by E. The susceptible, infected and recovered compartments are denoted by S, E and R respectively.

$$\frac{dS}{dt} = -\frac{\beta IS}{N} + \in R \tag{2.35}$$

$$\frac{dE}{dt} = \frac{\beta IS}{N} - \sigma E \tag{2.36}$$

$$\frac{dI}{dt} = \sigma E - \gamma I \tag{2.37}$$

$$\frac{dR}{dt} = \gamma I - \in R \tag{2.38}$$

The general functionality of the SEIR is also modelled by the ordinary different equations (ODE) approach, which puts into perspective all the four major compartments. This is denoted in equations 2.35 to 2.38 (Pandey et al., 2020). In these equations, β, σ, γ and $\in$ denote the infection rate, incubation rate, recovery rate, and susceptibility rate.

$$R_0 = \frac{\beta_0 \sigma}{(\mu+\alpha)(\mu+\gamma)} \tag{2.39}$$

The epidemic reproduction factor is then calculated from the four coefficients of the SEIR model as illustrated in equation 2.39 above (Pandey et al., 2020).

Susceptible, Symptomatic, Purely Asymptomatic, Hospitalized, Exposed, Recovered and Deceased (SIPHERD) model

This model was proposed by Mahajan et al. (2020) to predict COVID-19 spread. The main aim of this model was to integrate the asymptomatic COVID-19 cases and to use these cases to extract the underlying COVID-19 spread patterns and distributions. The SIPHERD model breaks down a population into seven distinct compartments of the pandemic. These are comprised of the susceptible, symptomatic, purely asymptomatic, quarantined, exposed, recovered, and deceased as explained by Mahajan et al. (2020).

$$\frac{dS}{dt} = -S(\alpha E + \beta I + \gamma P + \delta h) \tag{2.40}$$

$$\frac{dE}{dt} = S(\alpha E + \beta I + \gamma P + \delta h) - (\mu + \epsilon_I + \epsilon_P)E \tag{2.41}$$

$$\frac{dH}{dt} = \mu(E + P) + vI - \sigma H(t - t_R) - \tau H(t - t_D) \tag{2.42}$$

$$\frac{dI}{dt} = \in_I E - (v + \omega)I \tag{2.43}$$

$$\frac{dP}{dt} = \in_p E - (\mu + \eta)P \tag{2.44}$$

$$\frac{dR}{dt} = \omega I + \eta P + \sigma H(t - t_R) \tag{2.45}$$

$$\frac{dD}{dt} = \tau H(t - t_D) \tag{2.46}$$

The SIPHERD model can be derived using the ordinary differential equation methodology, which is a similar approach to that followed by the SIR and SEIR models described earlier. These ordinary differential equations for the SIPHERD model have been mathematically

depicted in equations 2.40 to 2.46 above (Mahajan et al., 2020). In these equations, S, E, I, P, H, R, and D represent the susceptible, exposed, symptomatic, purely asymptomatic, hospitalized, recovered, and deceased compartments, respectively. On the other hand, t_R and t_D denote the recovery and death lags, respectively (Mahajan et al., 2020).

Susceptible, Infected, Diagnosed, Ailing, Recognized, Threatened, Healed and Extinct (SIDARTHE) model

This model was proposed by Giordano et al. (2020) to predict the COVID-19 epidemic. It divides the population into 8 distinct compartments with regard to epidemic transmission. These compartments are made up of the susceptible, infected, diagnosed, ailing, recognized, threatened, healed, and extinct individual units.

$$S^{\cdot}(t) = -S(t)(\alpha I(t) + \beta D(t) + \gamma A(t) + \delta R(t)) \quad (2.47)$$

$$I(t) = S(t)(\alpha I(t) + \beta D(t) + \gamma A(t) + \delta R(t)) - (\in + \zeta + \lambda)I(t) \quad (2.48)$$

$$D^{\cdot}(t) = \in I(t) - (\eta + \rho)D(t) \quad (2.49)$$

$$A^{\cdot}(t) = \zeta I(t) - (\theta + \mu + \kappa)A(t) \quad (2.50)$$

$$R^{\cdot}(t) = \eta D(t) + \theta A(t) - (\nu + \xi)R(t) \quad (2.51)$$

$$T^{\cdot}(t) = \mu A(t) + \nu R(t) - (\sigma + \tau)T(t) \quad (2.52)$$

$$H^{\cdot}(t) = \lambda I(t) + \rho D(t) + \kappa A(t) + \xi R(t) + \sigma T(t) \quad (2.53)$$

$$E^{\cdot}(t) = \tau T(t) \quad (2.54)$$

The SIDARTHE model compartments can be mathematically computed as shown in equations 2.47 to 2.54 above (Giordano et al., 2020). In these differential equations, the transmission rate owing to encounters between a susceptible individual and an infected, diagnosed, ailing, recognized individual is denoted by α, β, γ, δ respectively. Generally, α is greater than γ, although this is likely to be greater than β and δ which would take place after taking into account the assumption that diagnosed people are in appropriate and efficient quarantine (Giordano et al., 2020). It is worth noting that precautions taken, such as social distancing strategies, can change such parameters. Another assumption taken when computing these parameters is that the contagion risk from vulnerable individuals treated in appropriate ICUs is thought to be minimal (Giordano et al., 2020).

θ and $\in$ represent the identification probability rate in asymptomatic and moderately symptomatic occurrences, correspondingly (Giordano et al., 2020). These variables, which are also adjustable, represent the degree of scrutiny paid to the illness and the frequency of examinations conducted on the population. Because a symptomatic patient is much more likely to be examined, θ is usually larger than $\in$. On the other hand, ζ and η signify the likelihood rate whereby an infected patient is unaware of becoming infected and also a context whereby the person is aware, respectively (Giordano et al., 2020).

The rates at which undiagnosed and recognized infected individuals acquire capital complications are represented by μ and ν, respectively (Giordano et al., 2020). τ signifies the rate of death. $\lambda, \kappa, \xi, \rho$ and σ signify the recovery rate for the 5 categories of infected people, and may change dramatically if an adequate illness therapy is discovered and applied to symptomatic patients (Giordano et al., 2020).

2.2. Deep learning prediction methods

In this section, the deep learning approaches applied widely in prediction problems are being discussed in detail. These deep learning methods are part of the broad group of machine learning approaches used in a wide range of problems, such as classification and clustering problems. Deep learning methods are unique in the way that they are largely dependent on artificial neural networks with a multi-representational perspective used in learning insights or patterns from data.

2.2.1. Convolutional neural network

The convolutional neural network (CNN) is one of the most popular deep learning algorithms, including 1D, 2D, and 3D CNN network architectures as described by T. Li et al. (2020). Sequence data processing, image and text identification, and medical image and video recognition are only some of the activities that CNNs are utilized for. CNN is a useful tool for time series analysis since it allows for the examination of patterns and deviations over time.

In addition, a CNN is made up of three layers: an input layer, many hidden layers, and an output layer. Activation functions, such as rectified linear units, may be included in each layer. Fully linked layers and processors incorporating convolutional layers, as well as

pooling, are common hidden layers. The goal of pooling is to provide invariance to smaller local deviations while reducing the feature space's complexity (Han et al., 2021). Coupled convolution and pooling operations are frequently repeated in the network in order to harvest comprehensive data. Convolutional solution enables for a much fewer set of variables than a fully linked network, allowing for more improved training and prediction as described by Han et al. (2021).

The CNN architecture consists of numerous layers that analyze inputs and provide an output. Convolutional layers and pooling layers, for example, are layers in a CNN model that fulfill different tasks, as explained by Xishuang Dong et al. (2017). This structure enables the model to learn many data properties, making it a useful tool for analyzing and forecasting outcomes. The convolutional layer is an important component of the CNN neural networks because it allows them to learn and process information efficiently. The parameters of the layer are a series of convolutional filters that may be tailored to have a narrow receptive element yet extend beyond the breadth of the raw data (T. Li et al., 2020).

The CNN's pooling component plays the most significant role during time series prediction since it performs the identification of short-term underlying patterns in a time series, which eventually provides more details concerning the relationship between the variables under study as described by K. Wang et al. (2019).

$$H_{ck} = f(W_{ck} {}^{*} X_{cT} + b_{ck}) \tag{2.55}$$

In equation 2.55, the final output vector can be obtained from the given parameters whereby, * represents the convolution component's computation, f denotes the activation function, X_{cT} represents the input data, W_{ck} and b_{ck} both denote the weight coefficients.

The pooling component acts as an integral part with the convolutional part to adjust the output vector scope after the filtration operations performed by the convolution components, which serves to eliminate the problem of overfitting as mathematically denoted in equation 2.55 (K. Wang et al., 2019). The CNN model has been widely applied alongside other methods when handling prediction problems. For example, it has been widely preferred because of its robustness in extracting inherent data structures such as patterns, trends, and seasonality elements, for example, in applications such as the prediction of gold prices by Livieris et al. (2020), where it was applied in the initial data processing phases.

2.2.2. Multi-layer perceptron network

A Multi-layer perceptron (MLP) model is a form of feedforward neural network that is comprised of one or several layers that exist between the input and output units. The term "feedforward" refers to data flowing in a single path from the source to the destination unit. As described by Victor Devadoss and Antony Alphonse Ligori (2013), MLP is composed of three main parts: an input unit, a hidden unit, and an output unit. The incoming data is transferred to the input unit's neurons, which process it according to their specific computation weights. The results obtained are subsequently passed to the hidden unit, and lastly to the output unit (Feng et al., 2020).

Weights are used to correlate neuro connections, and adjusting the weights in a particular way enables the connected model to learn. The learning or training process is the method through which the weights in the network are changed (Feng et al., 2020). One of the most widely used learning approaches is the backpropagation methodology. A forward pass and a backward pass are used in this method (Victor Devadoss & Antony Alphonnse Ligori, 2013).

In the forward pass, an input sequence is transmitted to the network's units, with the outcome being a pair of expected values at the output units with fixed weight values. The standard error is computed in the backward pass through drawing a distinction seen between the model's actual result and the desirable outcome supplied to the model, and it is conveyed backward throughout. The weights are modified in this step to bring the network's actual result closer to the anticipated result.

For a given vector of input and output values, the multi-layer perceptron algorithm extracts the nonlinear approximation that exists in the data between the input and output sets as illustrated in equation 2.56 to equation 2.59 below (Feng et al., 2020).

$$X = x_1, x_2, \dots, x_R \tag{2.56}$$

In equation 2.56, X denotes the input vector with R number of data points.

$$y = y_1, y_2, \dots, y_S \tag{2.57}$$

In equation 2.57, y represents the output set that consists of S number of data points.

$$f(.) : X \rightarrow y \tag{2.58}$$

The MLP model in equation 2.58 above then aims to determine the nonlinear functional correlation between the input and output vectors represented in equations 2.56 and 2.57 above.

$$Y = g\left(\sum_{i=1}^{R} w_{ji} x_i + b_j\right) \tag{2.59}$$

The final output of the MLP algorithm can be obtained mathematically from equation 2.59 above. Y represents the final output at a particular node j in this equation. The nonlinear function is illustrated by g. w_{ji} denotes the weight value while $b_j$illustrates the bias coefficient.

2.2.3. Auto-Encoder model

This is a deep learning method that is based on the neural network functionality that is used in prediction problems (L. Wang et al., 2017). The Autoencoder algorithm archives its objectives by incorporating an encoding and decoding compartment into the general architecture. The autoencoder performs the significant task of encoding the input vectors to another vector called the hidden representation vector using a compact set of hidden units which comprise of activation functions (L. Wang et al., 2017).

$$h_i = f_\theta(x_i) \tag{2.60}$$

For a particular set of input values x_i, the set of feature values can be obtained from equation 2.60 above, where f_θ denotes the encoder unit and h_i represents the feature set of values (Han et al., 2021).

$$r_i = g_\theta(h_i) \tag{2.61}$$

The decoder unit of the autoencoder model is used to construct feature mappings to the input vector in a backward format using a parameterized function as illustrated in equation 2.61 above (Han et al., 2021). In this equation, the decoder transforming function is denoted by g_θ.

$$f_\theta(x_i) = \sigma_f(Wx + b) \tag{2.62}$$

In equation 2.62, the encoder function is expanded and its inherent parameters and operation have been mathematically illustrated (Han et al., 2021). The weight factor has been represented by W. The bias is depicted by b while the activation function is denoted by σ_f.

$$g_\theta(h_i) = \sigma_g(\, W'h + \, b') \tag{2.63}$$

The decoder operation of the autoencoder model can be mathematically expanded and represented in equation 2.63 above (Han et al., 2021). In this equation, the bias, weight, and activation function terms are denoted by characters b', W' and σ_g respectively.

Using the stochastic gradient descent method or backpropagation techniques, the autoencoder model's parameters can be effectively determined by focusing on the total reduction of the error factor observed when performing the reconstruction of the output vector values (Han et al., 2021), (L. Wang et al., 2017). Variations of the autoencoder algorithm have been proposed. These include methods such as the sparse autoencoder, regularized autoencoder encompassing the contractive autoencoder, and denoising autoencoder as indicated by Han et al. (2021).

2.2.4. Restricted Boltzmann machine

The restricted Boltzmann machine (RBM) is a distinct kind of Boltzmann machine model which differs from the basic Boltzmann machine model by not incorporating the hidden and visible units into the complete structure of the network. The Boltzmann machine algorithm is a stochastic generative deep learning method that can establish a probabilistic model throughout its inputs by including the hidden and visible units of the neural network, as explained by Zhang et al. (2015).

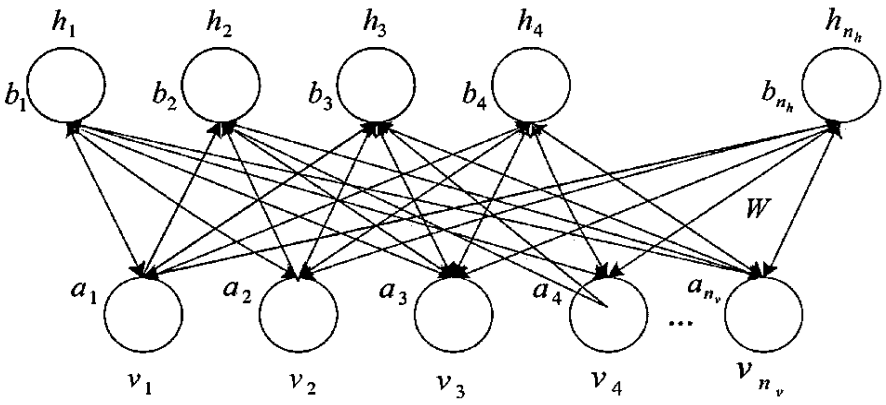

Figure 2.4. Illustration of the Restricted Boltzmann machine model.

One level of binary-valued observable neurons and another level of Boolean hidden nodes make up the RBM architecture overall. As illustrated in figure 2.4 above, there are no bidirectional and symmetrical interconnections across neurons in the same level inside a RBM model, yet complete bidirectional and symmetrical interconnections occur across neurons in separate levels as described by H. Wang et al. (2016). In figure 2.4, the hidden and visible nodes are represented by the characters, h and v respectively. Their respective biases are denoted by b and a. The main energy function of the RBM model can be derived using these parameters, including the weight parameter, which is denoted by W as illustrated in equation 2.32 (Han et al., 2021).

$$E(v,h) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{ij} W_{ij} v_i h_j \tag{2.64}$$

In equation 2.64 above, the a_i and b_j denote bias values for the visible and hidden layers, respectively. v_i and h_j denote the visible and hidden layer nodes, respectively, while the weight is depicted by W_{ij}.

$$P(v,h) = e^{-E(v,h)}/Z \tag{2.65}$$

The probability distribution that exists across the hidden and visible levels is computed from the energy function as illustrated in equation 2.65 above as described by H. Wang et al. (2016). In this equation, Z denotes the partition function used to ensure that there is a normal distribution (Han et al., 2021). After determining the conditional probabilities of each node given its layer, using the activation functions, the unknown model parameters are obtained in unsupervised format for better prediction results as described by H. Wang et al. (2016).

2.2.5. Deep belief network

The main building blocks of the deep belief network (DBN) comprise the RBM model described in section 2.2.4 above. The DBN model has been successfully applied to a wide range of prediction problems, such as the prediction of gold prices conducted by P. Zhang and Ci (2020).

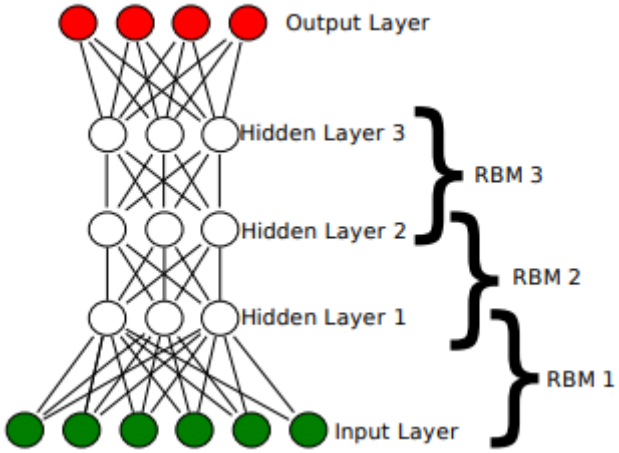

Figure 2.5. Illustration of the DBN model architecture.

As illustrated in figure 2.5 above, DBN models are created by hierarchically stacking RBM models as described by Hamel and Eck (2010) and H. Wang et al. (2016). The first level, in particular, has been pre-trained as a discrete RBM using training data for inputs. The output of the layer located on the level beneath is often used as an input to the hidden units once the properties of the first hidden layer are needed (P. Zhang and Ci, 2020). These 2 hidden levels are therefore treated as a new RBM and trained accordingly. Lastly, to deploy the DBN system, a conventional classifier, namely logistic regression, is introduced to the outermost layers and trained in a supervised setting as indicated as described by H. Wang et al. (2016). The optimization process, also known as gradient descent, aims to incrementally alter the properties of the whole system.

$$p(v) = \sum_h p(h|\theta)p(v|h,\theta) \tag{2.66}$$

In a DBN model, the final visible vector probability is calculated using the model parameters that have been learned by the RBM model, which includes the inherent probabilities across the hidden vectors. In equation 2.66 above, the probability of obtaining a visible vector $p(v)$ can be mathematically computed as described by Mohamed et al. (2009). In this equation, θ represents the model parameters obtained from the trained RBM model operations. $p(v|h,\theta)$ is retained after learning, whereas $p(h|\theta)$ can always be substituted with a stronger model trained by employing the hidden layers created from the training phase as input samples for the next RBM model. Under the new combination, this substitution establishes a variational lower constraint on the likelihood of the training set (Mohamed et al., 2009).

2.2.6. Generative Adversarial Nets (GAN)

GAN is a strong group of generative models that work by intuitively modeling complex data. Its main idea is based on the two-player zero-sum system, where both players' overall returns are zero and then each person's gains or losses of value are perfectly balanced by the losses and gains of value of another player (K. Wang et al., 2017). GANs are frequently made up of a generator and a discriminator that are both trained at the same time. The generator creates unique data while attempting to capture the possible distribution of genuine samples.

A binary classifier is frequently used as the discriminator, separating genuine samples from created samples as precisely as feasible (Y. Wang et al., 2020). GANs use a minimax game optimization method, with the objective of reaching Nash equilibrium, when the generator is thought to have approximated the pattern of real samples as described by K. Wang et al. (2017).

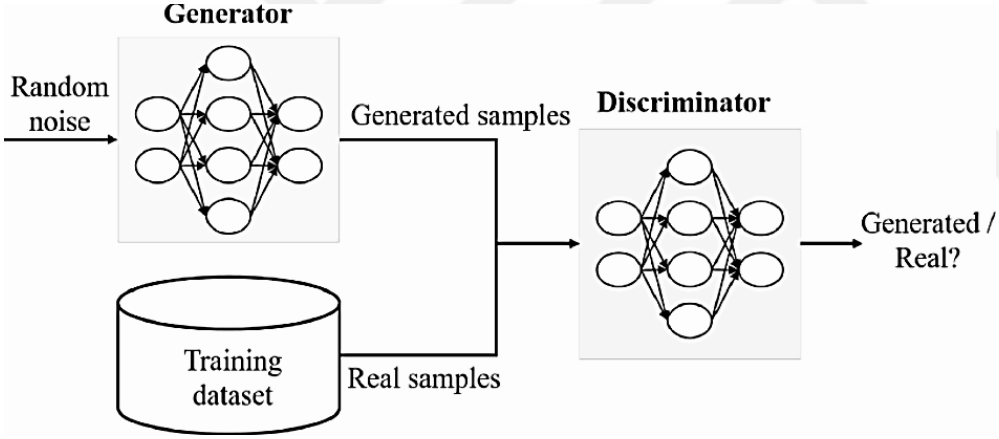

Figure 2.6. Illustration of the GAN architecture.

In figure 2.6 above, the general architecture of the GAN model has been given (Y. Wang et al., 2020). It is clear in this figure that the generator neural network performs the task of random generation of input samples that are drowned out by a noise latent space. On the other hand, it is evident that the discriminator neural network acts as a filter in the middle of the input from the true observations and artificially produced inputs from the generator neural network.

$$L_D = -E_{S_r}\left[log\left(D(S_r;\theta_D)\right)\right] - E_Z\left[log\left(1 - D\left(S_g;\theta_D\right)\right)\right] \quad (2.67)$$

The loss function of the discriminator neural can be mathematically obtained from equation 2.67 above (Y. Wang et al., 2020). The anticipated value across all real observations and the anticipated value across all arbitrary inputs to the generator neural network, respectively, are indicated by E_{S_r} and E_Z in this equation. The parameters of the discriminator neural network are labeled as θ_D, while the actual and synthesized samples are indicated as S_r and S_g, respectively. The discriminator function is denoted by D.

$$L_G = -E_Z\left[log\left(1 - D(G(z;\theta_G);\theta_D)\right)\right] \quad (2.68)$$

The loss function of the generator neural network can be obtained mathematically from equation 2.68 above (Y. Wang et al., 2020). In this equation, G denotes the generator function while θ_G represents the generator neural network model's parameters. The noise vector is labelled by z.

2.2.7. Deep Stacking Network

The Deep Stacking Network (DSN) is an extensible deep learning model that can train in parallel to extract crucial model parameter coefficients (Hutchinson et al., 2013). It learns using the context of a supervised system, block-by-block manner, without the necessity for stochastic gradient descent across all units, as is the case with many neural networks (Han et al., 2021). The DSN blocks are layered to build the entire deep learning model, each consisting of a basic component. Every DSN block is a perception unit that has just one hidden level that has been adjusted (Hutchinson et al., 2013). A higher level weight vector connects the input and hidden units, while the weights for the level beneath connect the logistic sigmoidal nonlinear hidden layer to the linear output layer.

$$U = (HH^T)^{-1}HT^T = F(W)\ (H = [h_{\mathbf{1}}, \dots, h_i, \dots, h_N]\ ,\ h_i = \sigma(W^T x_i) \quad (2.69)$$

In equation 2.69 above, the weights of the different layer nodes have been mathematically denoted (Han et al., 2021). In this equation, the hidden layer blocks are denoted by H while the upper layer units have been illustrated by U. The weight of the lower layer nodes is denoted by W. x_i denotes the training set while T indicates the target vectors.

It is important to note that each unit of the typical DSN produces an approximation of the ultimate label class, which is merged with the original input matrix to generate the enlarged "input" matrix for the following block of the DSN (Hutchinson et al., 2013). Since each

successive block keeps the original data, it will always perform more effectively on the training dataset than the prior block. The foundation of the DSN is significantly streamlined, and enhancing connection weights is inherently computationally efficient due to the stringent essence of building hierarchies as well as the clarity of every block (Han et al., 2021).

It should also be noted that the label class approximation vector, is continuously weighted and is calculated as a weighted sum of the hidden units within every block. The label is determined by the index of the maximum value in the output vector, which is determined at the top block. The output vector is used to combine with the input space matrix to transmit to the DSN's subsequent upper unit in the bottom blocks of the DSN (Hutchinson et al., 2013).

2.2.8. The Extreme Learning Machine

ELM is a single-hidden layer feedforward network (SLFN) with arbitrarily constructed input weights and biases and empirically determined output weights (Wan et al., 2014). The key concept underlying ELM is to reduce challenging nonlinear optimization problems, such as determining the ideal set of weights, hidden node biases, and output weights, to a simple least square problem of identifying the appropriate output weights (X. Chen et al., 2012). It indicates that users may not have to take into account every one of the input weights and hidden layer biases as long as the weights' norm is modest enough, while the output weights seem to be the only aspect to take note of. This concept differs significantly from traditional recurrent learning approaches (X. Chen et al., 2012).

$$f(x_j; w, b, \beta) = \sum_{i=1}^{k} \beta_i g\left(w_i x_j + b_i\right), j = 1, \dots, N \quad (2.70)$$

The ELM model can be mathematically computed using equation 2.70 above (Wan et al., 2014). In this equation, w_i denotes the weight set that links the *i*th hidden units with the source units. The weight that links the output units with the *i*th hidden units is denoted by β_i. The threshold of the *i*th is represented by b_i. g represents the activation function. On the other hand, the number of hidden units is depicted by k.

ELM has been proven to be conceptually aimed at fulfilling unified estimation, as well as having high generalization performance and exceptionally rapid efficiency. This can be observed from its methodology that the clients' sole task is to choose the activation function

and the hidden units' figure, making it simple to operate as described by X. Chen et al. (2012). The ELM avoids several of the drawbacks of traditional gradient-based training methods, such as local minima, overtraining, and excessive computational costs, by avoiding repetitive gradient-based training. The ELM with neurons in the hidden layer can retrieve unique samples perfectly without error for almost any indefinitely discrete activation function. Furthermore, according to the supplied input weights, ELM training can still offer the greatest outcomes. Because of the simple matrix computation, the training rate is exceptionally rapid. (Wan et al., 2014).

2.2.9. Prophet model

The Prophet model is a time series prediction deep learning approach. As an open-source project, the Facebook group designed and promotes this approach. It is based on the generic definition of a generative additive model (GAM), which is a rectilinear regression strategy whereby the linear parameter is dependent on smoothing operations, according to S. J. Taylor and Letham (2017). Equation 2.71 can be used to illustrate GAMs empirically.

$$g\big(E(Y)\big) = \beta_0 + \ f_1(x_1) + \ f_2(x_2) + \cdots + \ f_m(x_m) \tag{2.71}$$

Y stands for the univariate response variable, x_1 stands for the predictor variable, and f_1 stands for the smoothing coefficients in equation 2.71. The Prophet approach provides a range of advantages because of its use of the GAM modeling process, including adaptability and short fitting periods, and assesses a time series challenge from three viewpoints, encompassing trend, seasonality, and holiday elements, as stated in a study conducted by S. J. Taylor and Letham (2017). The trend aspect considers whether time series data is likely to increase or decrease with time. Seasonality, on the other hand, examines data variations over a brief span of time.

$$g(t) = \frac{C}{1 + exp(-k(t - m))} \tag{2.72}$$

The Prophet model's trend element may be further divided into two main sections. These components, according to S. J. Taylor and Letham (2017), are comprised of the rating growth model and the piece linear model. The saturating growth model, the logistic growth criteria is utilized, that is formally expressed in equation 2.72. C stands for carrying capacity, k for growth rate, and m for offset. Carrying capacity and growth rates are not always consistent. A time-based dynamic capacity $C(t)$ is used in such situations. The first step is

to define the change point j, which is formally specified in equation 2.73. Adjustments in the rate of growth, on the other hand, are dealt with by creating a number of variables involved within which fluctuations in the rate of growth are allowed.

$$\gamma_{\mathrm{j}} = \left(s_j - m - \sum_{l<j} \gamma l\right)\left(1 - \frac{\mathrm{k} + \sum_{l<j} \delta_l}{\mathrm{k} + \sum_{l<j} \delta_l}\right) \tag{2.73}$$

In equation 2.73, the base rate that occurs at an arbitral time t is denoted by k. On the other hand, at a particular time s_j, the growth rate that exists is denoted by δ_l.

$$g(t) = \frac{C(t)}{1 + exp(-k(+ a(t)^T \delta)(t - (m + a(t)^T \gamma)))} \tag{2.74}$$

Upon putting into consideration the possibilities of fluctuations in the rate of expansion and carrying capacity, the final mathematical expression of the logistic growth model is shown in equation 2.74 above (S. J. Taylor & Letham, 2017).

$$s(t) = \sum_{n=1}^{N} \left(a_n cos \frac{2\pi \mathrm{nt}}{\mathrm{P}}\right)\left(b_n sin \frac{2\pi \mathrm{nt}}{\mathrm{P}}\right) \tag{2.75}$$

As illustrated in equation 2.75 above, seasonality is handled in the Prophet framework by using the Fourier series (S. J. Taylor & Letham, 2017). In this equation, the definite time span of the time series is indicated by p.

$$Z(t) = [1(t \in Di), \ldots, 1(t \in DL)] \tag{2.76}$$

As can be seen in equation 2.76, holidays and events are first modeled in the Prophet model using only a vector of regression coefficients with the premise that they too are independent (S. J. Taylor & Letham, 2017). Di denotes a collection of upcoming and past days for a certain holiday i.

$$h(t) = Z(t)\kappa \tag{2.77}$$

Equation 2.77 yields the ultimate holiday model, assuming that every holiday i is allocated a property k_i that is immediately correlated to the prediction function (S. J. Taylor & Letham, 2017).

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \tag{2.78}$$

As illustrated in equation 2.78 above, the ultimate projected result $y(t)$ is considered to consist of the trend, seasonal, and holiday constituent functions, whereby ε_t reflects changes not recorded by the algorithm (S. J. Taylor & Letham, 2017).

3. FORECASTING THE SPREAD OF COVID-19 USING DEEP LEARNING AND BIG DATA ANALYTICS METHODS

In this study, both the statistical and deep learning models were selected to perform the prediction of the COVID-19 pandemic in African countries on a regional basis. In section 2, the statistical and deep learning strategies commonly used for prediction both in the health sector and generally in other fields have been given in detail. Among the statistical models are models ranging from the simplest linear regression models to advanced autoregressive models such as the GARCH and ARIMA models. Other statistical models used in prediction problems include nonlinear regression and exponential smoothing models. Nonlinear regression models play a critical role in forecasting problems when there exists no significant linear relationship between the dependent and independent variables used in a prediction problem.

The exponential smoothing models have several variations depending on the way they solve the prediction problems, whether in an additive or multiplicative format. On the other hand, the ARIMA models combine the predictive strength of both the autoregressive and moving average models with a difference factor. Deep learning methods make use of extensive layers of artificial neural networks to learn patterns that may exist in the data. This makes such approaches more robust compared to some of the statistical methods, which excel only when certain conditions are met, such as for ARIMA models that require data to be stationary. Among the commonly applied deep learning methods are the CNN, MLP, autoencoder models, deep belief models, general adversarial nets, extreme learning machines, Prophet models, and recurrent neural networks such as the LSTM and gated recurrent units (GRUs) models.

3.1. Recurrent neural networks (RNNs)

Conventional feedforward algorithms have proven to be effective in a variety of domains. Information flow adjustments are transmitted in one path across hidden units in these kinds of models, with the outcome controlled solely by the present condition (Shen et al., 2018). Such neural network models, on the other hand, have little storage capacity and are therefore unsuitable for defining input sequences and time correlations in past data (Zeroual et al., 2020). Recurrent neural networks were designed to solve time-dependent learning

difficulties in order to overcome this constraint. The primary idea behind RNNs is to take into account the impact of previous data when constructing the result (Salehinejad et al., 2017). To accomplish this, units representing gates that influence the result based on past data are placed in the result (Zeroual et al., 2020). The RNN structure allows RNNs to retain, recall, and analyze complicated data from the past over longer timeframes (Salehinejad et al., 2017).

The basic structure of the RNN model is comprised of 3 core layers. These layers include the input, recurrent hidden and the output layers. For a given input layer with a set of N units, the input data units consist of a set of vectors that serve as the input units to the input layer. That is to say, if for a given time frame t, the input layer units consist of a vector of values $\{\dots, x_{t-1}, x_t, x_{t+1}, \dots\}$ whereby in this vector, $x_t = (x_1, x_2, \dots, x_N)$ (Salehinejad et al., 2017). The input layer units are directly connected to the hidden layer, which is also connected to the output layer.

$$y_t = f_O(W_{HO}h_t + b_o) \tag{3.1}$$

After computations of the RNN model, the final output can be received from the output layer and can be mathematically derived as represented in equation 3.1 above (Salehinejad et al., 2017). In this equation, y_t represents the output vector while f_O denote the activation function. W_{HO} represents the weight and h_t depicts the hidden layer unit vector. The bias of this model is denoted by b_o.

3.2. Gated recurrent units (GRUs)

The GRU model is made up of only two major gates, which consist of the update gate and the reset gate in its structure (Zeroual et al., 2020) (Shen et al., 2018). The update gate takes on the role of ensuring a stable amount of memory for prior computations. On the other hand, the reset gate maintains the link that exists between the new source data and the prior memory (Zeroual et al., 2020).

$$r_t = s(W_r[h_{t-1}, x_t]) \tag{3.2}$$

The reset gate manages the effect that the output from the memory cells from a prior point in time has on the existing data value (Shen et al., 2018). In a circumstance whereby this value is not so relevant, the reset gate is opened to ensure that the memory's output value's

effect on the current value is nullified. The operation of the reset gate is depicted in equation 3.2 above (Shen et al., 2018). In this equation, x_t denotes the input vector while h_{t-1} denotes the memory cell output value. The weight of the reset gate is depicted by W_r.

$$u_t = s(W_u[h_{t-1}, x_t]) \tag{3.3}$$

The update gate plays an important role in the GRU model. It deals with the task of determining whether to neglect the existing data or not. The operating mechanism of the update gate also has the advantage of significantly reducing the gradient vanishing challenge faced in most neural network models (Shen et al., 2018). The operation of the update gate is mathematically denoted in equation 3.3 above (Shen et al., 2018). The weight of this gate is indicated by W_u.

3.3. Long short-term memory (LSTM)

According to a study conducted by Zoabi et al. (2021), the LSTM model is a deep learning algorithm which is based on the recurrent neural network framework. The LSTM model was intended to address the limitations of the conventional recurrent neural network by bridging time lag gaps while lowering inaccuracies by utilizing less computational power. Three essential fundamental elements form the LSTM network: The forget, input, and output gates are among them (Shastri et al., 2020). The forget gate determines how much information from history is lost. The input gate gets signals that are fed into the cell's internal state, whereas the output gate generates every next hidden state from the current internal state data.

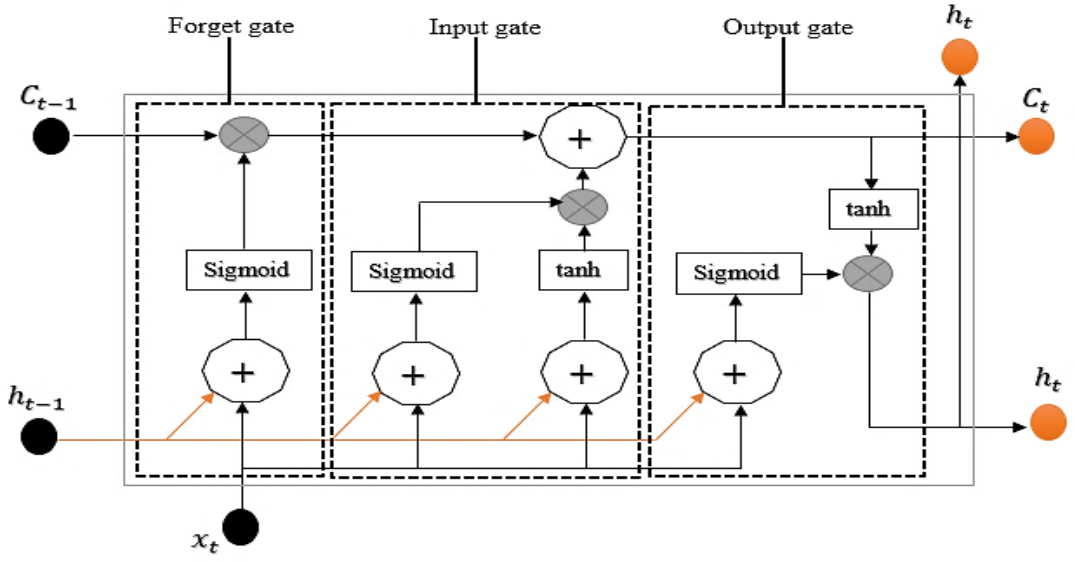

Figure 3.1. Illustration of the LSTM model architecture.

The three main building blocks of the LSTM model have been illustrated in the figure 3.1 above. In this illustration, the present cell state and hidden states have been denoted by C_{t-1} and h_{t-1} respectively. The input and output vectors have been denoted by x_t and h_t respectively.

The forget gates are used to refresh the internal state of every unit. The LSTM model begins to work whenever an input signal is received by a cell that will then be computed to generate an output at a particular time frame.

$$f_t = \sigma W f(h_{t-1}, x_t) \tag{3.4}$$

The forget algorithm accepts the outcome and decides whether to keep or reset the data. As shown in a study conducted by Kırbaş et al. (2020), this particular operation can also be mathematically denoted in equation 3.4. The sigmoid function is computed to produce a number that indicates whether the data in a specific cell should be retained or not, whereupon the final result is acquired.

$$i_t = \sigma\, Wf(h_{t-1}, x_t\ ) \tag{3.5}$$

$$\tilde{C}_t = tanh[Wc\ (h_{t-1}, x_t\ )\ ] \tag{3.6}$$

$$\tilde{C}_t = f_t * \tilde{C}_{t-1} + i_t * \tilde{C}_t \tag{3.7}$$

$$o_t = \sigma[Wo(h_{t-1}, x_t\ )] \tag{3.8}$$

$$h_t = o_t * tanh(C_t) \tag{3.9}$$

The LSTM model operations can be expressed mathematically in the equations 3.5-3.9 above (Kırbaş et al. (2020). In these equaions, W and σ denote the weight and the activation function respectively. On the other hand, f_t signifies the forget function, C_t indicated the feature set, whereas o_t signifies the outcome of the activation function.

3.4. Thesis

In this section, the data sources, methods, and tools used in this study are described in detail.

3.4.1. Data source

In this study, countries from the five major regions of the African continent were studied. These regions included the Northern, Southern, Central, Eastern, and Western regions. The COVID-19 dataset used was obtained from the Humanitarian Data Exchange open-source site (Africa: Covid-19 Infections (National)-Humanitarian Data Exchange, 2021). This data is comprised of the COVID-19 cumulative cases for all African countries in an ungrouped format. This data was grouped based on the five African regions. The data regarding the population of the various countries from the 5 regions was obtained from the open source worldometer online platform (African Countries by Population (2021) - Worldometer, 2021). This source compiles population demographic data that is expounded by an assortment of international organizations such as the United Nations.

3.4.2. Geographical regions and populations of Africa

Nations from the 5 main regions of the African continent were utilized as case studies throughout this research. The Northern, Eastern, Southern, Central, and Western divisions of the African continent are illustrated in figure 3.2 below, based on the work compiled by Wikipedia contributors (2022b).

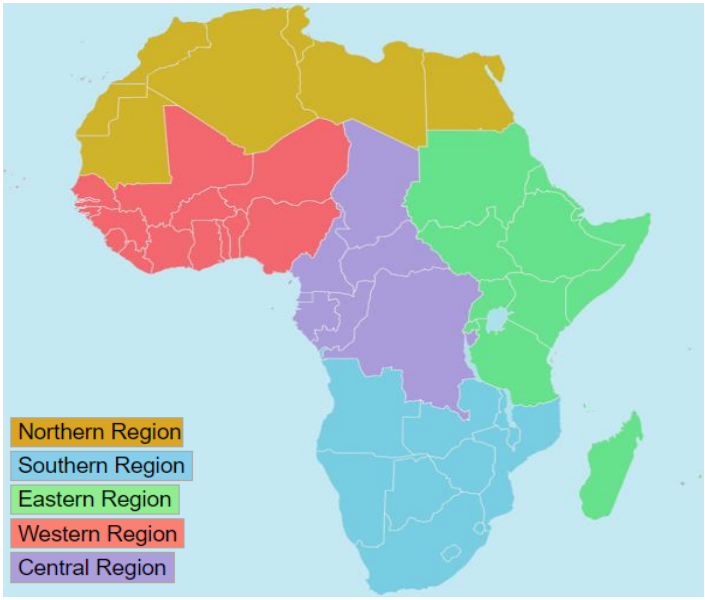

Figure 3.2. The five primary regions of the African continent

In figure 3.2 above, the 5 broad regions of the African continent have been clearly represented. The countries studied in this research can be tracked based on their respective regions, as displayed in the figure above. A comparative analysis is later performed based on each region to determine the riskiest nation for eventual precautions for the region as a whole.

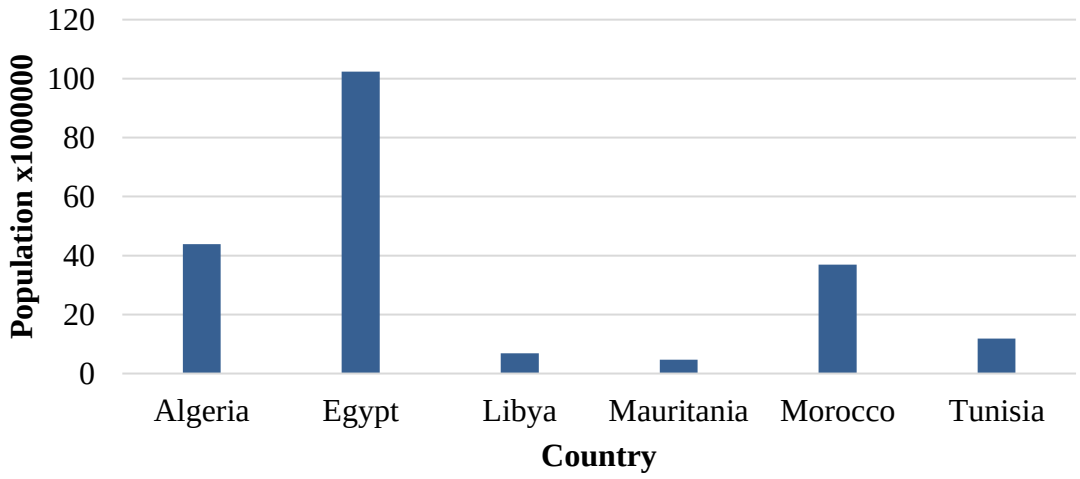

Figure 3.3. Illustration of the Northern African nations' populations

As can be seen in figure 3.3, Egypt is by far the most heavily inhabited nation in the Northern African region. In contrast, Mauritania possesses the smallest total population. The population density in each of the five nations apart from Egypt is just under 60 million.

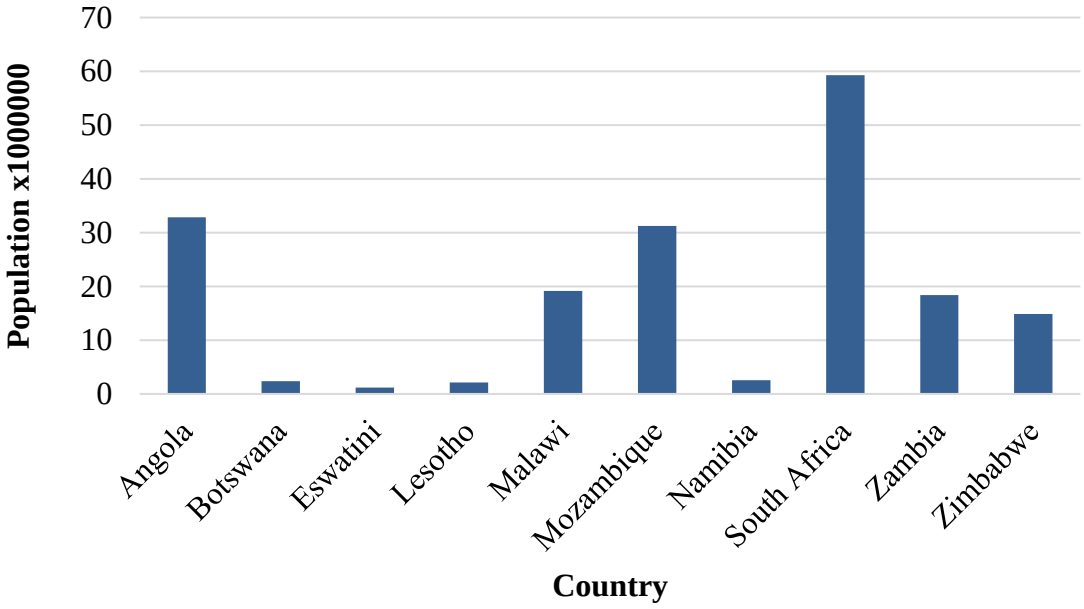

Figure 3.4. Illustration of the Southern African nations' populations

In this study, 10 Southern African nations were researched, and the populations of these nations are depicted in Figure 3.4. South Africa has the highest population density inside this region, whereas Eswatini has the lowest population density. Over 50% of the states have populations of under 20 million.

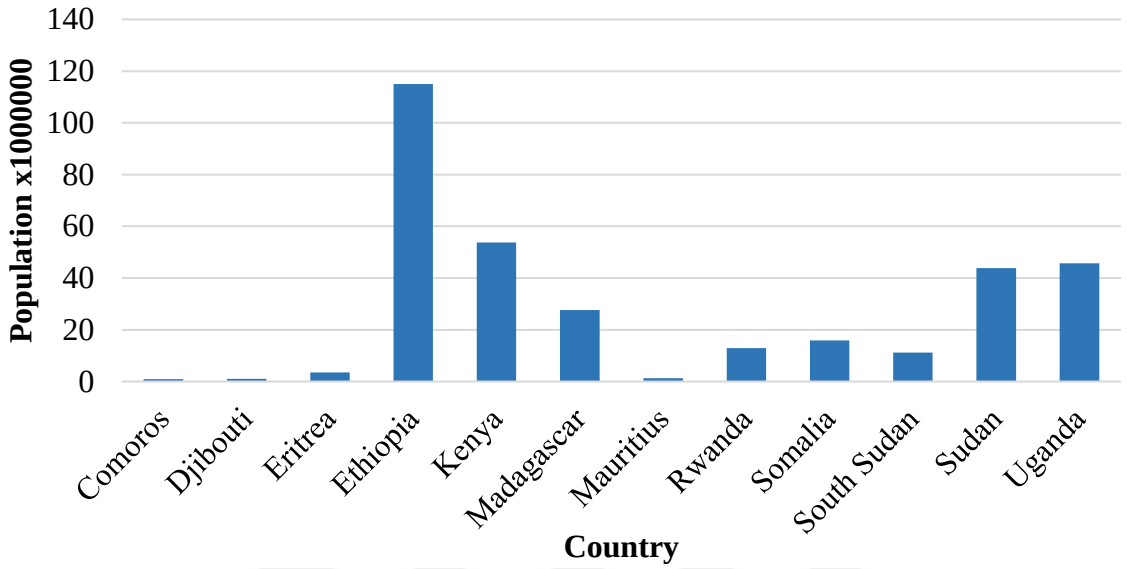

Figure 3.5. Illustration of the Eastern African nations' populations.

With a population of almost 100 million inhabitants, Ethiopia is indeed the most populated nation as depicted in figure 3.5. With the exception of Ethiopia, the remainder of the nations are fairly inhabited, with populations significantly lower than 60 million.

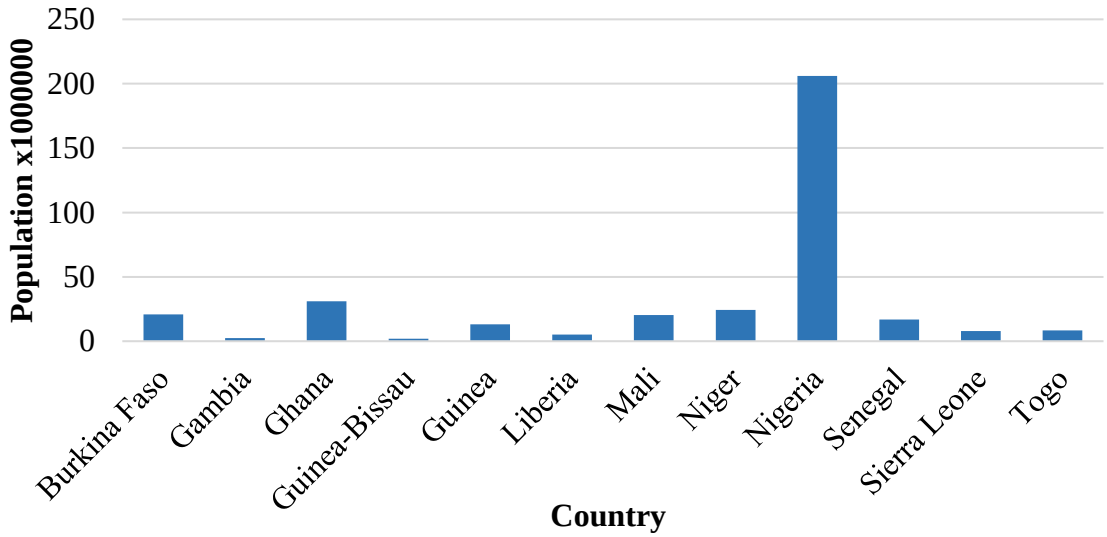

Figure 3.6. Illustration of the Western African nations' populations

The populations of the Western region are depicted in Figure 3.6, and where most states, with the exception of Nigeria, possess populations of less than 50 million inhabitants. Nigeria has a much larger population than the other countries in the region, with approximately 200 million inhabitants. The situation of the COVID-19 outbreak in this nation is expected to have a substantial impact on its spread across the region.

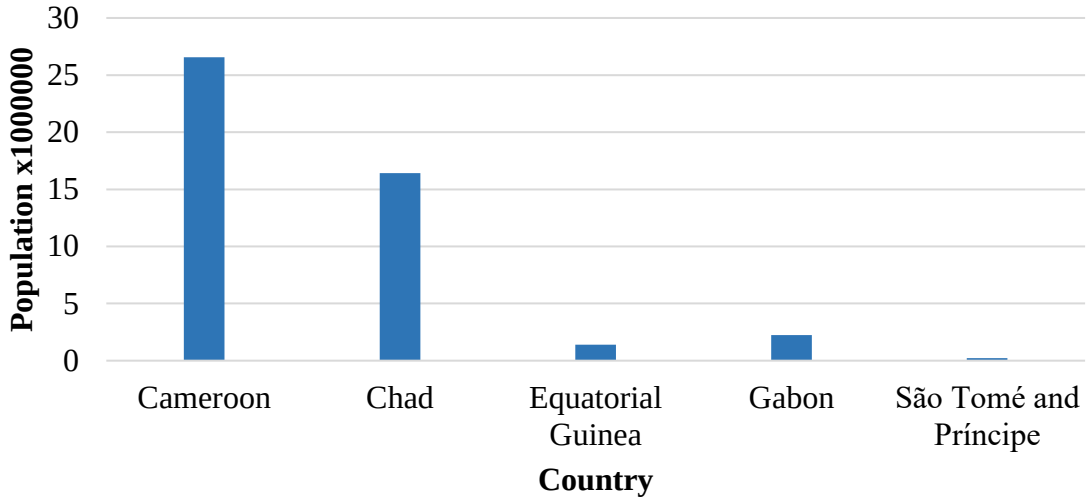

Figure 3.7. Illustration of the Central African nations' populations.

Cameroon, the most populated nation in the Central African region, has a population of less than 30 thousand individuals. The least densely populated nation in this region is observed to be São Tomé and Príncipe as shown in the Figure 3.7.

3.4.3. Justification for the implemented models.

In this study, three prediction algorithms were chosen for the task of predicting the spread of COVID-19 on the African continent. In this section, the main reasons why these three models were opted for are described in detail. These three models include the ARIMA, LSTM, and Prophet models. In the earlier sections 2.1.6, 2.2.9, and 3.3, the mathematical operating principles of the ARIMA, Prophet, and LSTM models have been elucidated in a comprehensive manner.

LSTM

The LSTM model is a subset of RNN models that aims to eliminate the shortcomings of classical models by capturing the inherent patterns, seasons, and trends that exist within a particular input sequence, as indicated in a study conducted by Yu et al. (2021). Since the LSTM algorithm contains memory units which operate mostly as links across recent and

historical pieces of data, it is able to provide preferable results. Essential information units with significant targeted observations are preserved, whereas pieces of information with lower weights are discarded with the long short-term memory model's forget component. The forget functionality streamlines the algorithm to focus on capturing the dependencies that emerge in a given set of inputs, as well as improving the overall accuracy by removing noisy elements from the learnt information during this step.

The LSTM algorithm, as defined by Zeroual et al. (2020), provides a solution to the challenge of vanishing gradients that disproportionately affect classic RNN algorithms, in which the generated gradient varies between extreme thresholds, especially when performing model training. According to a study conducted by Arbel (2021), the LSTM algorithm comprises of activation vectors that are employed in the forget gate component to compute the gradients, which enables it to address the commonly faced challenge of vanishing gradient. The batch and epoch values, for instance, are easily adjustable hyperparameters of the LSTM algorithm. The availability of these model-tuning features further makes the LSTM model easier to tune and streamline for best results (Yu et al., 2021).

ARIMA

According to Singh et al. (2020) research, the ARIMA model is a statistical methodology that employs regression to correlate historic data units and errors by employing weight variables for better accuracy. This model combines the advantages of both the autoregression and moving average approaches, making it a reliable option for prediction tasks. It's a versatile model to use since it takes into account the differences across a sequence both from a historical and current perspective, allowing it to accommodate and compute non-stationary data with just a smaller number of input variables, as detailed by Abdulmajeed et al. (2020).

Additionally, the availability of tuning approaches such as the PACF and ACF graphs, among others, as stated in a research study done by Gebretensae and Asmelash (2021), makes it simpler to identify the best configuration coefficients of the ARIMA model. Furthermore, criteria like the Akaike information criteria and Bayesian information criteria can be used to assess how efficient an ARIMA forecast is with a particular set of hyperparameter coefficients, which makes it easy to optimize the predictive performance. According to Y. Wang et al. (2012), the ARIMA algorithm can analyze the data containing seasons and trend patterns by enhancing the hyperparameter coefficients to accommodate

these patterns. The above allows any seasonal association across the COVID-19 dataset to be captured.

Prophet

The Prophet model is an additive regression deep learning technique developed by Facebook that includes a functional structure, as shown by Abdulmajeed et al. (2020), that focuses on various seasonal fluctuations inside a particular data series, which includes annual, weekly, and daily trends. The Prophet algorithm has proven to perform effectively on sequences that comprise incomplete data items and outliers, as described by Y. Wang et al. (2012). Therefore, this makes it an appropriate alternative for processing and predicting COVID-19 data, which contains these factors.

The Prophet model, as shown by Letham and Taylor (2017), has modeled predictive support, which manages non-linear curves whenever the natural limit is attained, and versatility in optimization, which includes techniques such as smoothing capabilities that identify and compute the seasonality factors in the data to make an excellent match influenced by historical periods. With the Prophet model, it is very simple to collect and derive the impacts of activities such as vacations in series data (Letham, Taylor, 2017). Because of all these characteristics, this model is a good fit for predicting the spread of the COVID-19 pandemic.

3.4.4 The research architecture

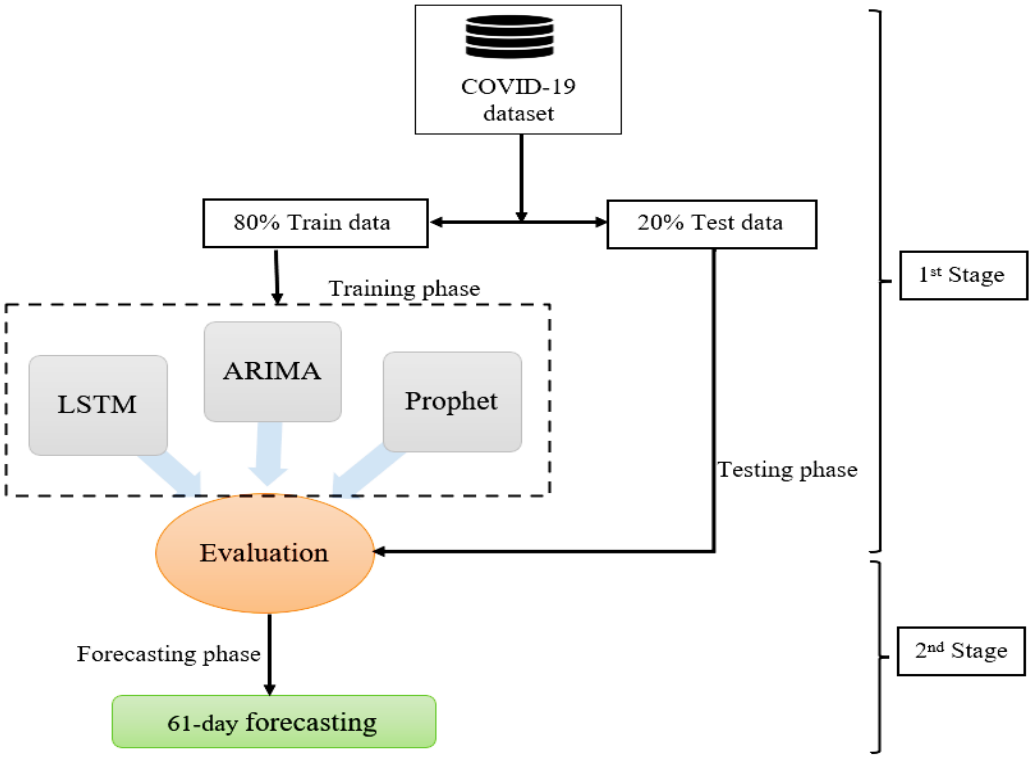

Figure 3.8. Illustration of the methodology applied in this study.

The significant phases of this research are depicted in Figure 3.8, incorporating dividing the precompiled COVID-19 cumulative case data into 80 percent training and the remaining 20 percent as testing datasets. The algorithms were then fitted using the training data. The models' predictive accuracy was determined using seven performance criteria metrics. Lastly, the best functioning and efficient algorithm was selected to forecast the COVID-19 instances for another 61 days ahead of time.

3.4.5 Model performance metrics

In this study, seven metrics were adopted to assess the predictive performance of the models. These metrics include, the Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), Root Mean-Square Error (RMSE), Symmetric Mean Absolute Percentage Error (SMAPE), Mean Absolute Percentage Error (MAPE), Normalized Root Mean-Square Error (NRMSE), and R2 score.

Mean Square Error

This is a quantitative indicator for assessing a prediction model's accuracy. It calculates the average score for both the squares of the expected and observed variables. Following equation 3.21 (Kırbaş, 2020), the mean squared error can be calculated numerically.

$$MSE = \frac{1}{n}\sum_{1=1}^{n}(\hat{Y}_i - Y_i)^2 \quad (3.10)$$

The overall number of observations n, the exact value Y, and the anticipated value $\hat{Y}$ are all represented in equation 3.10. The Euclidean distance requirements are used in this performance indicator, which means its score is mostly expected to be a non-negative value.

Root Mean Square Error

The RMSE is a model selection statistic that is employed to assess the effectiveness of the predictive algorithm. It integrates the errors of the estimated variable into a unified predictive performance criterion. The RMSE can be calculated using equation 3.11.

$$RMSE = \sqrt{\frac{1\sum_{1=1}^{n}(\hat{Y}_i - Y_i)^2}{n}} \quad (3.11)$$

The overall number of observations n, the actual value Y, and the anticipated value $\hat{Y}$ are all represented in equation 3.11. The square root of the exponential average of both the expected and observed vectors is used to calculate this metric.

Mean Absolute Percentage Error

This metric uses a percentage figure to evaluate the reliability of a learning algorithm. It produces a non-scaled accuracy score by calculating the mean within the absolute percentage errors in the projected values (de Myttenaere et al., 2016). Equation 3.12 can be used to represent this performance measure numerically.

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \quad (3.12)$$

The observed vector of numbers is represented by A_t, the projected value is expressed by F_t, and the total number of data points is represented by n in equation 3.12. The MAPE statistic is simple to grasp and apply as a method for evaluating predictive accuracy because the ultimate result is a percentage value. This statistic overcomes the disadvantage of positive

and negative values canceling out because absolute values are utilized ("MEAN ABSOLUTE PERCENTAGE ERROR (MAPE)," 2000).

Symmetric Mean Absolute Percentage Error

The SMAPE metric is a relative error-based model performance indicator. Equation 3.13 can be used to represent this measurement numerically.

$$SMAPE = \frac{100\%}{n}\sum_{t=1}^{n}\frac{|F_t - A_t|}{(|A_t| + |F_t|)/2} \quad (3.13)$$

The observed vector numbers are represented by A_t, the forecasted value is represented by F_t, and the overall number of observations is represented by n in equation 3.13.

Peak Signal-to-Noise Ratio

This statistic lets you compute and determine the signal-to-noise ratio whenever the signal is at its most extreme peak (Kourav et al., 2018).

$$PSNR = 20log_{10}(\frac{MAX_f}{\sqrt{MSE}}) \quad (3.14)$$

The highest signal value is expressed by MAX_f in equation 3.14. MSE stands for mean square error.

Normalized Root-Mean-Square Error

This performance measure is used to obtain a complete assessment of prediction accuracy that uses the RMSE metric value. It eliminates the difficulties that variable scales present in both the information that is computed by the algorithms and the particular algorithms.

$$NRMSE = \frac{RMSD}{Y_{max} - Y_{min}} \quad (3.15)$$

The root-mean-square deviation (RMSD) is defined in equation 3.15. The RMSD measure is also known as the RMSE statistic.

R2 score

The R2 score is also commonly known as the coefficient of determination. It is employed to determine the amount of variance between the dependent variable and the independent variable.

$$R2 = \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - Y^i)^2} \quad (3.16)$$

In equation 3.16, the projected values are represented by f_i, whereas the original values are represented by y_i, and the mean is represented by Y^i.

4. RESULTS AND DISCUSSION

Nations from the African continent have been used in this research. These nations have been divided into five categories, as depicted in earlier sections. The ARIMA, LSTM, and Prophet algorithms were among the forecasting methods employed. The model performance accuracy values generated from these algorithms are presented in this section of every African region.

4.1. Model training and evaluation

4.1.1. Northern Africa

In this section, the COVID-19 cumulative cases are comparatively studied, and the model performance in the six nations of this region is discussed in detail.

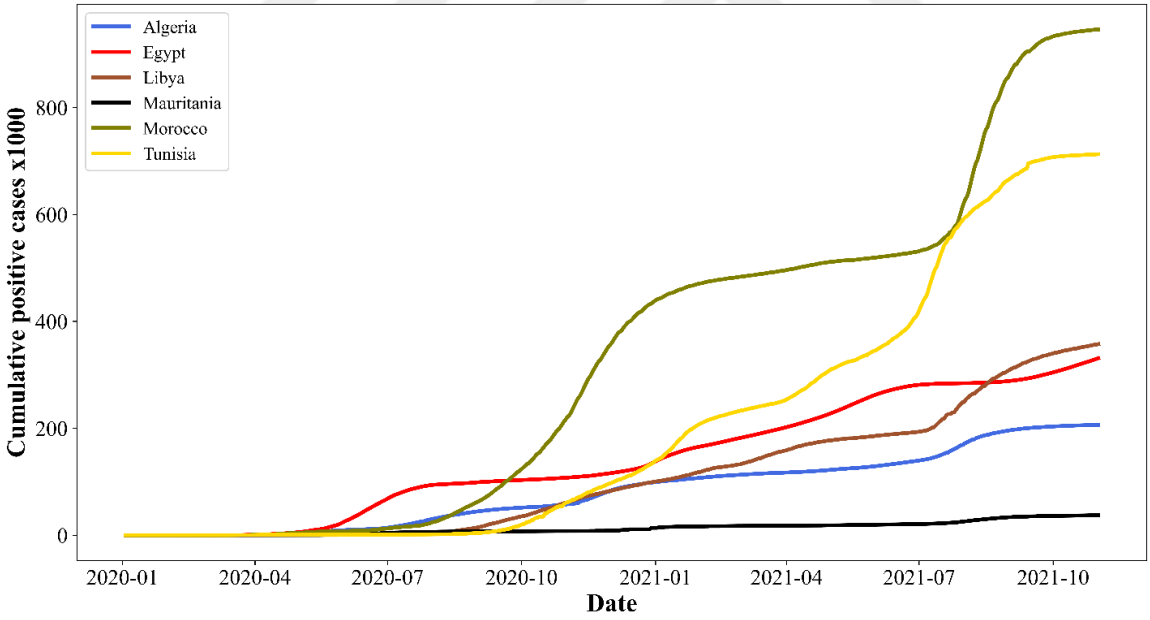

Figure 4.1. Cumulative positive cases in the Northern African region

As shown in Figure 4.1, Morocco has sustained the highest prevalence of COVID-19 cases. Tunisia was next in line in this dire situation. Mauritania, on the contrary, has had the fewest incidences throughout the whole period in comparison to the other countries in the same area. The number of cases in Libya is increasing, with a slight increase from October 2020 through July 2021. Well beyond the month of July, there is a significant spike that

gradually decreases near the month of October. The initial wave of COVID-19 instances in Libya fit this description well. Algeria is on the same path as Libya.

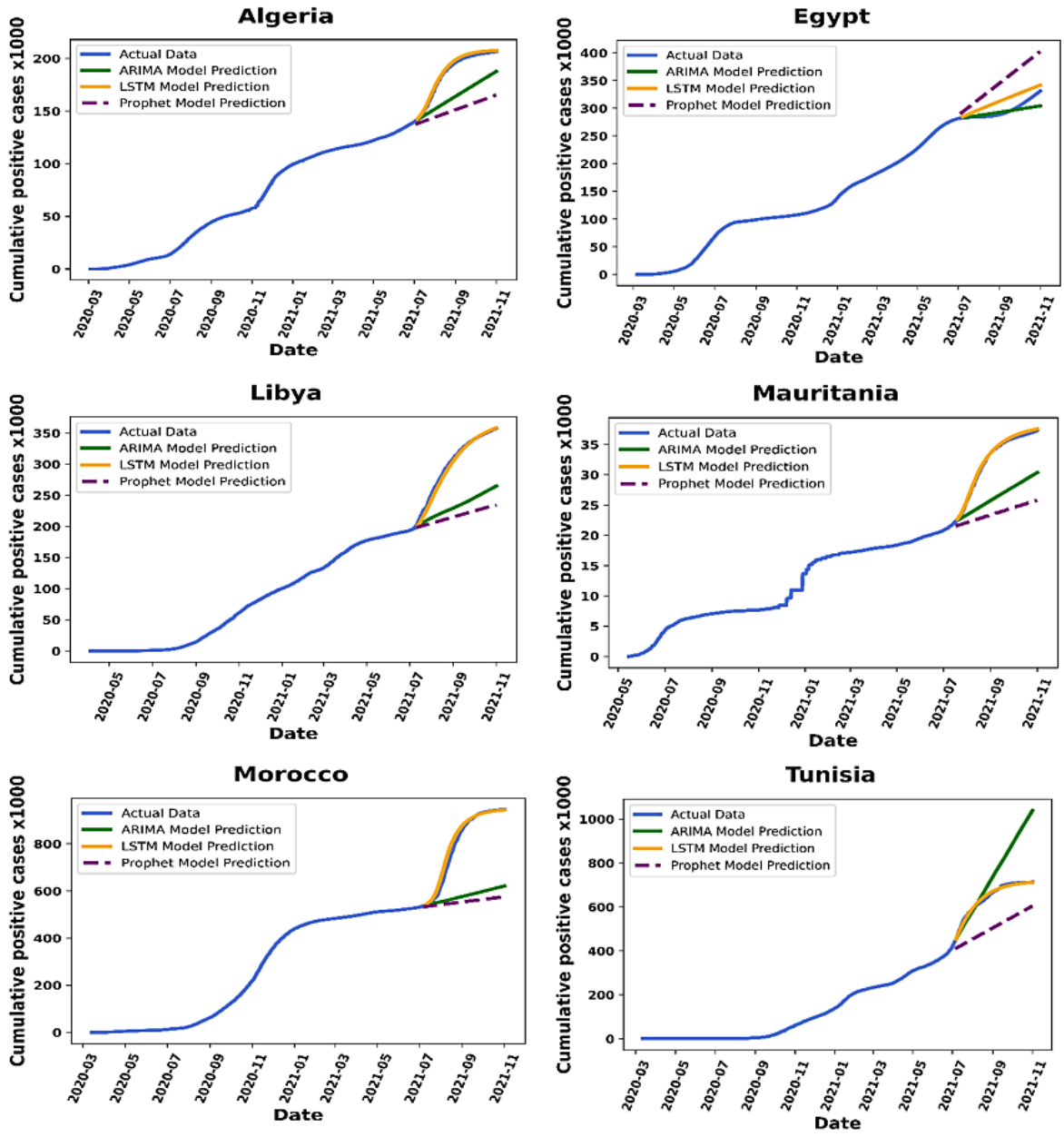

Figure 4.2. Actual and predicted cumulative positive cases in Northern Africa

As shown by figure 4.2, the LSTM algorithm performs better than the rest of the models. The Prophet model achieves the least in Tunisia. This is because the Prophet model forecasts a sudden spike of more than 800,000 instances, whereas the test data curve flattens. With respect to the observed data in nations like Egypt and Tunisia, the ARIMA and Prophet models projected smaller and larger case numbers, respectively. Besides these two nations, both algorithms anticipated a smaller number of cases based on the actual data in the remaining 4 nations. This reinforces the poor performance of these two algorithms when

contrasted to the LSTM model, which, with the exception of Egypt, predicts better outcomes that are closer to the actual data in five nations.

Table 4.1. Model performance in Northern Africa

Country	Model	PSNR	R value	NRMSE	SMAPE	RMSE	MSE	MAPE
	ARIMA	-28.3210	-1.0387	0.4392	20.0932	6646.4399	44175163.3400	18.0172
Mauritania	LSTM	-1.0514	0.9962	0.0190	0.7515	287.8118	82835.6322	0.7551
	Prophet	-31.3765	-3.1201	0.6244	30.2207	9448.5330	89274775.8500	25.8536
	ARIMA	-39.6473	-0.3443	0.3715	12.4041	24485.1867	599524367.7000	11.5554
Algeria	LSTM	-18.1055	0.9906	0.0311	1.0187	2050.2803	4203649.3090	1.0244
	Prophet	-43.4308	-2.2125	0.5744	20.2395	37851.3220	1432722577.0000	18.1621
	ARIMA	-59.9838	-1.7886	0.6184	29.8376	254525.4081	64783183368.0000	25.0037
Morocco	LSTM	-38.0839	0.9820	0.0497	1.8985	20452.0670	418287044.6000	1.9400
	Prophet	-60.8397	-2.3961	0.6824	33.7894	280882.9106	78895209467.0000	27.7835
	ARIMA	-49.3128	-1.4477	0.4721	24.5338	74504.0709	5550856581.0000	21.5143
Libya	LSTM	-25.4671	0.9899	0.0303	1.4573	4785.1562	22897719.8600	1.4395
	Prophet	-51.1119	-2.7040	0.5808	31.1302	91650.8128	8399871487.0000	26.4318
	ARIMA	-31.1439	0.5917	0.1873	2.0405	9198.8946	84619661.8600	2.0052
Egypt	LSTM	-37.1818	-0.6398	0.3753	5.5535	18434.4519	339829016.9000	5.7408
	Prophet	-46.5591	-13.2078	1.1048	15.3217	54261.6346	2944324989.0000	16.7849
	ARIMA	-55.2048	-2.9832	0.5456	13.7406	146819.1010	21555848418.0000	15.5706
Tunisia	LSTM	-28.8865	0.9907	0.0264	0.9365	7093.5494	50318443.0900	0.9371
	Prophet	-54.7576	-2.5934	0.5182	23.6105	139450.1419	19446342076.0000	21.0428

The best outcome based on the PSNR and R metrics can be found in table 4.1, with larger values indicating that the algorithm comparatively achieves better results. In Mauritania, the LSTM algorithm generated the largest PSNR and R scores. -1.0514 and 0.9962 were the scores obtained, correspondingly. With the exception of PSNR and R scores, the best outcomes are seen with lower numbers for the remaining evaluation metrics. In Mauritania, the LSTM model produced the smallest RMSE of 287.8118. MAPE values of 2.0052–25.0037 and 16.7849–26.4318 were obtained using the ARIMA and Prophet algorithms.

The LSTM algorithm, on the contrary, had the smallest MAPE (0.7551–5.7408). The LSTM algorithm's maximum MAPE score is notably less than that of the ARIMA and Prophet algorithms' lowest MAPE scores. Within the Northern African region, the LSTM algorithm is the best method for projecting the COVID-19 cases. Mauritania had the highest predictive accuracy in this region, whereas Morocco had the least, with a Prophet algorithm attaining a RMSE score of 280882.9106

4.1.2. Central Africa

Five nations were investigated in this region. At the time of conducting this research, Cameroon, with a population density of 26545863, was by far the most populous nation in this region. The least densely populated nation, on the contrary, is São Tomé and Príncipe, with a total population density of 219,159 inhabitants.

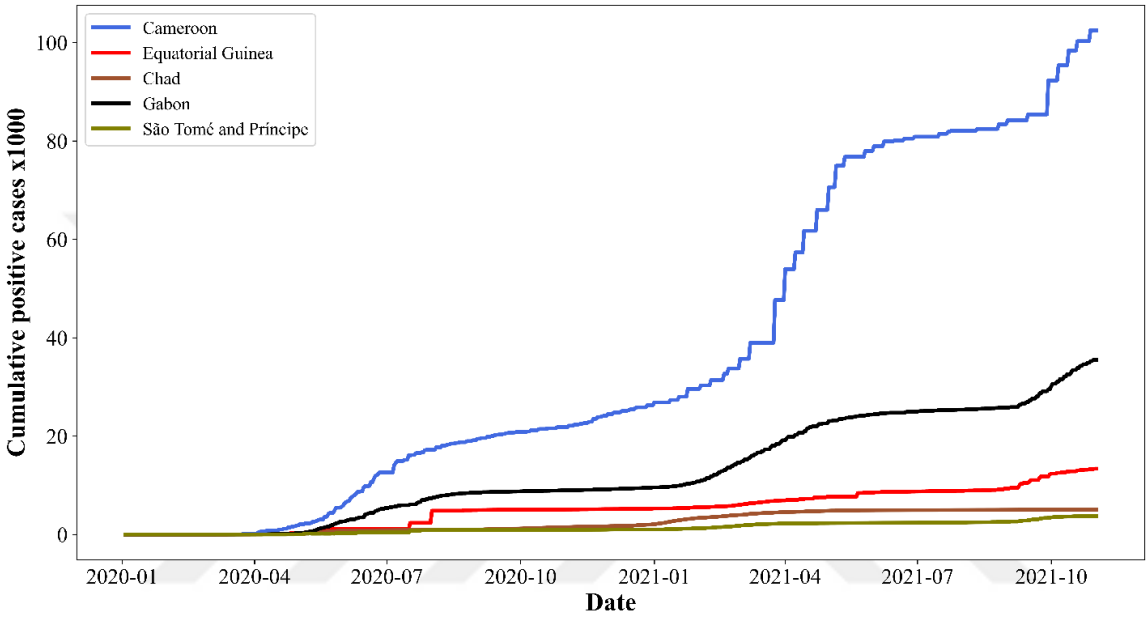

Figure 4.3. Cumulative positive cases in the Central African region

The COVID-19 cases obtained from the 5 nations in this region are shown in Figure 4.3. As shown by this figure, COVID-19 incidents in Cameroon are significantly greater than the average of the rest of the nations, with much more than two substantial waves. Cameroon is preceded by Gabon, which has 2 significant waves as well. The remaining nations follow a similar pattern, with moderate rises in COVID-19 instances.

São Tomé and Príncipe has the smallest number of instances. In both the population factor and the COVID-19 case number, there is a notable, strong correlation to be examined. This is due to the fact that Cameroon, the most populous country in the region, has the highest prevalence of COVID-19 incidents. However, the least populous nations, São Tomé and Príncipe, have the lowest instances. As a result, Cameroon is indeed the country in this region with the greatest potential risk of COVID-19 dissemination.

Table 4.2. Model Performance in Central Africa

Country	Model	PSNR	R value	NRMSE	SMAPE	RMSE	MSE	MAPE
	ARIMA	-23.7735	0.6842	0.1819	3.8398	3937.4568	15503566.0519	3.8655
Cameroon	LSTM	-29.9249	-0.3019	0.3694	7.0175	7994.4080	63910559.2705	7.4279
	Prophet	-40.3369	-13.3144	1.2249	24.9969	26508.4573	702698308.4259	28.8557
	ARIMA	-22.9030	-0.0850	0.3428	7.2343	3561.9664	12687604.6347	6.6077
Gabon	LSTM	-5.0523	0.9822	0.0439	1.0980	456.2018	208120.0823	1.1101
	Prophet	-19.1268	0.5452	0.2219	7.4839	2306.1026	5318109.2017	7.7485
	ARIMA	14.9747	-0.3830	0.3921	0.7312	45.4783	2068.2758	0.7272
Chad	LSTM	25.6706	0.8822	0.1144	0.2213	13.2742	176.2044	0.2212
	Prophet	0.0557	-41.9264	2.1842	4.6249	253.3710	64196.8636	4.7480
	ARIMA	-19.4805	-0.9611	0.5230	16.0832	2401.9573	5769398.8710	13.9253
Equatorial Guinea	LSTM	-10.6582	0.7428	0.1894	5.8417	869.8616	756659.2032	6.1844
	Prophet	-12.1101	0.6407	0.2238	8.8312	1028.1208	1057032.3790	8.8613
	ARIMA	-7.0908	-0.2709	0.4364	12.6397	576.8751	332784.8810	11.1389
São Tomé and Príncipe	LSTM	16.3165	0.9942	0.0295	0.8522	38.9686	1518.5518	0.8449
	Prophet	-8.3251	-0.6886	0.5030	15.7833	664.9582	442169.4077	13.7283

According to the performance metrics in Table 4.2, the LSTM algorithm in Chad gave the smallest RMSE score of 13.2742, while the Prophet algorithm in Cameroon provided the largest RMSE score of 26508.4573. Both the Prophet and LSTM algorithms, correspondingly, reported the smallest and greatest PSNR scores of 40.3369 and 16.3165 for Cameroon and São Tomé and Príncipe. The LSTM algorithm had the greatest MAPE span of 0.2212–7.4279, preceded by the ARIMA and Prophet algorithms with 0.7272–1.1389 and 4.7480–28.8557 correspondingly. The LSTM algorithm had the highest predictive accuracy, whereas the Prophet algorithm had the lowest.

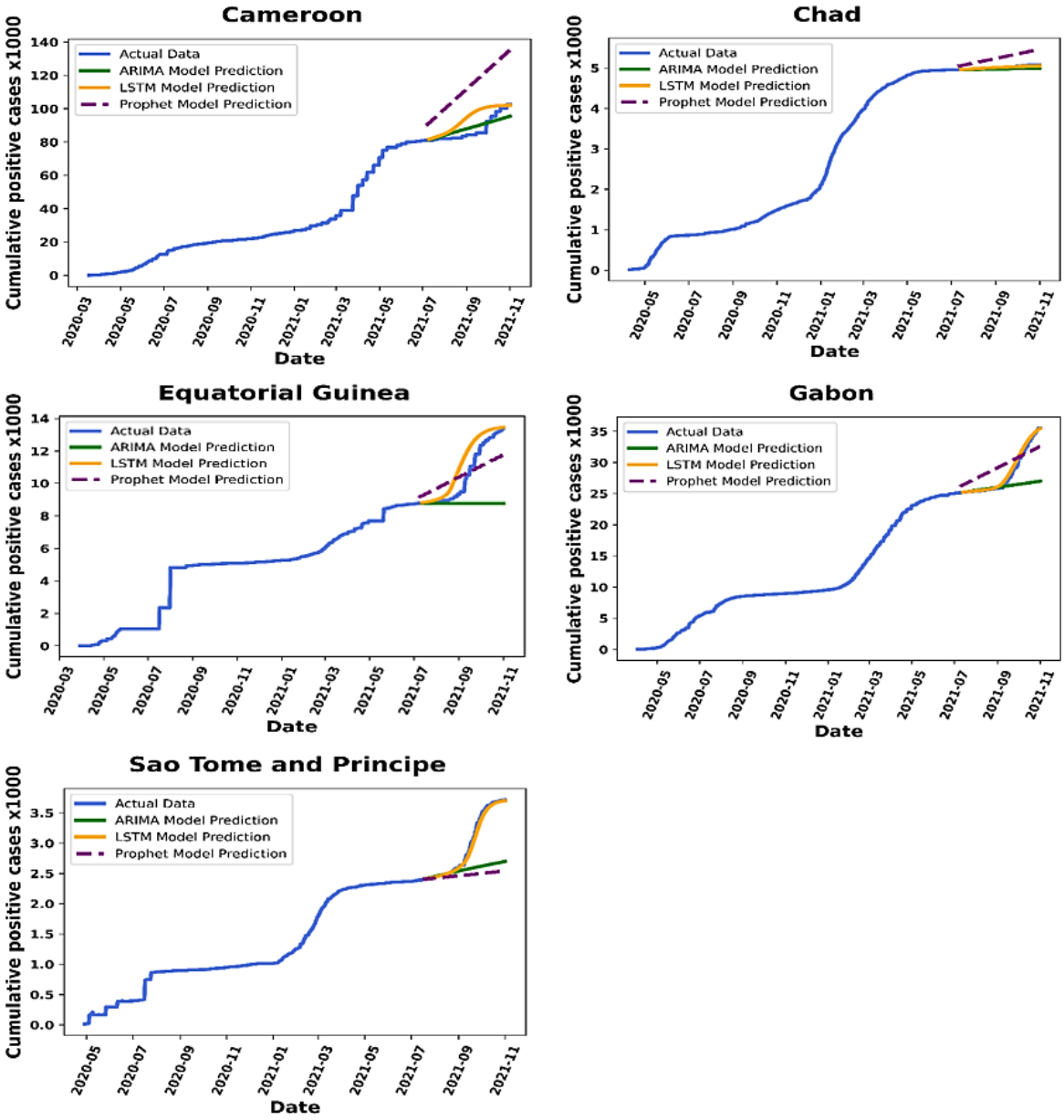

Figure 4.4. Actual and predicted cumulative positive cases in Central Africa

Figure 4.4 depicts a representation of the predictive accuracy of the models after test data prediction in several Central African nations. The LSTM algorithm's estimate substantially fits the actual data in 3 nations. This indicates that the LSTM algorithm produced the best results in this area. The Prophet algorithm, for instance, in Cameroon, is likewise shown to have the lowest model predictive accuracy. The ARIMA algorithm predicts significantly well in Chad, but it emerges in second place in the rest of the nation’s well after the LSTM algorithm.

4.1.3. Southern Africa

This research included ten nations from this region. South Africa, with a population density of 59308690, is the most populous nation in the region. On the other hand, Eswatini, with a population density value of 1160164, is the least inhabited nation in this region.

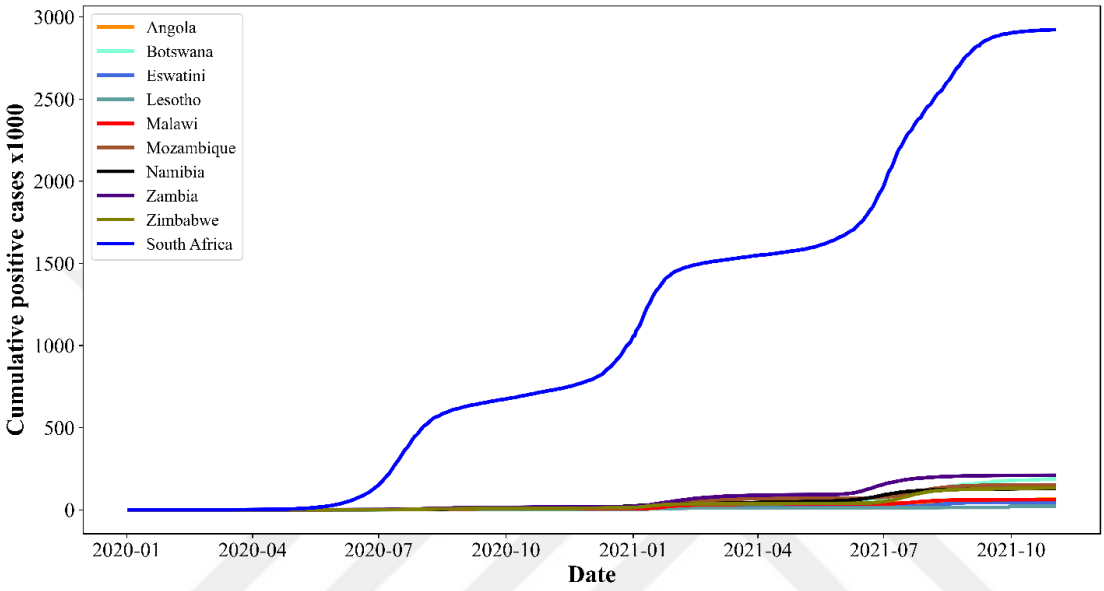

Figure 4.5. Cases in the Southern African region including South Africa

Figure 4.5 shows that, when contrasted to certain other nations in the region, South Africa does have the largest number of cases. This demonstrates how quickly the COVID-19 virus spreads across the nation. This puts adjacent nations in the same region at a significant risk of enhanced virus propagation. While the rest of the region is undergoing its second wave of virus transmission, South Africa is witnessing three episodes. There is a positive association between the huge number of cases recorded and the dense population value because it has the biggest population.

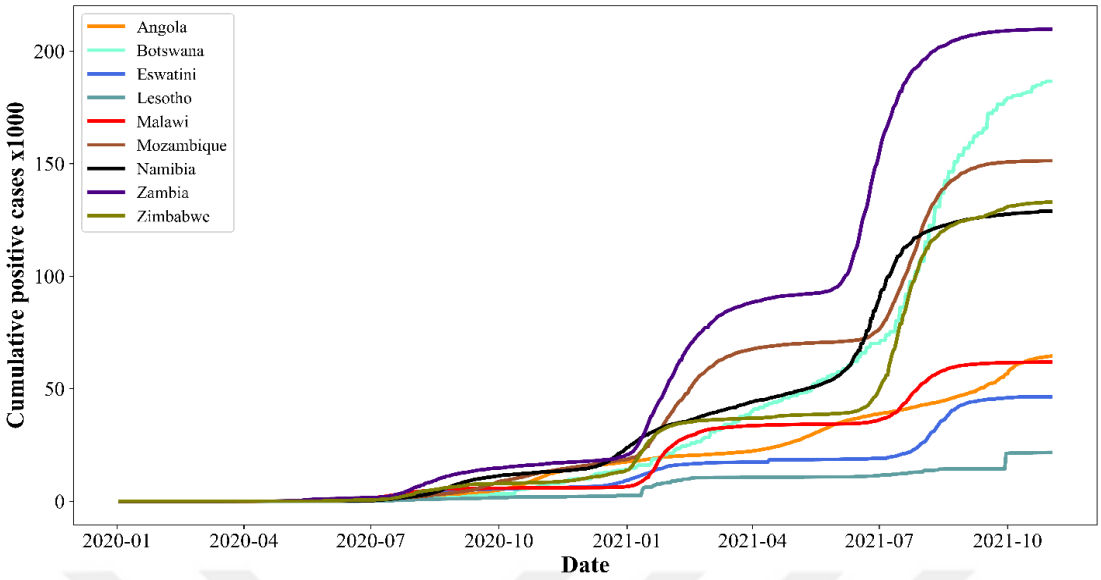

Figure 4.6. Cases in the Southern African region, excluding South Africa.

In order to do a comprehensive evaluation of the COVID-19 condition in other nations in the same region, South Africa was eliminated from Figure 4.6 for brevity. With the exception of South Africa, Zambia does have the highest concentration of cases when compared to the other nations. This is also the first nation where the number of incidents has increased. All nations have likewise seen their second strongest surge of COVID-19 transmission. It's worth mentioning that Lesotho had the smallest number of instances. After October, a consistent number of incidents with flatter trends could be seen in all nations. This perfectly indicates the impact of prevention measures and immunizations, as well as other methods of controlling the spread of disease.

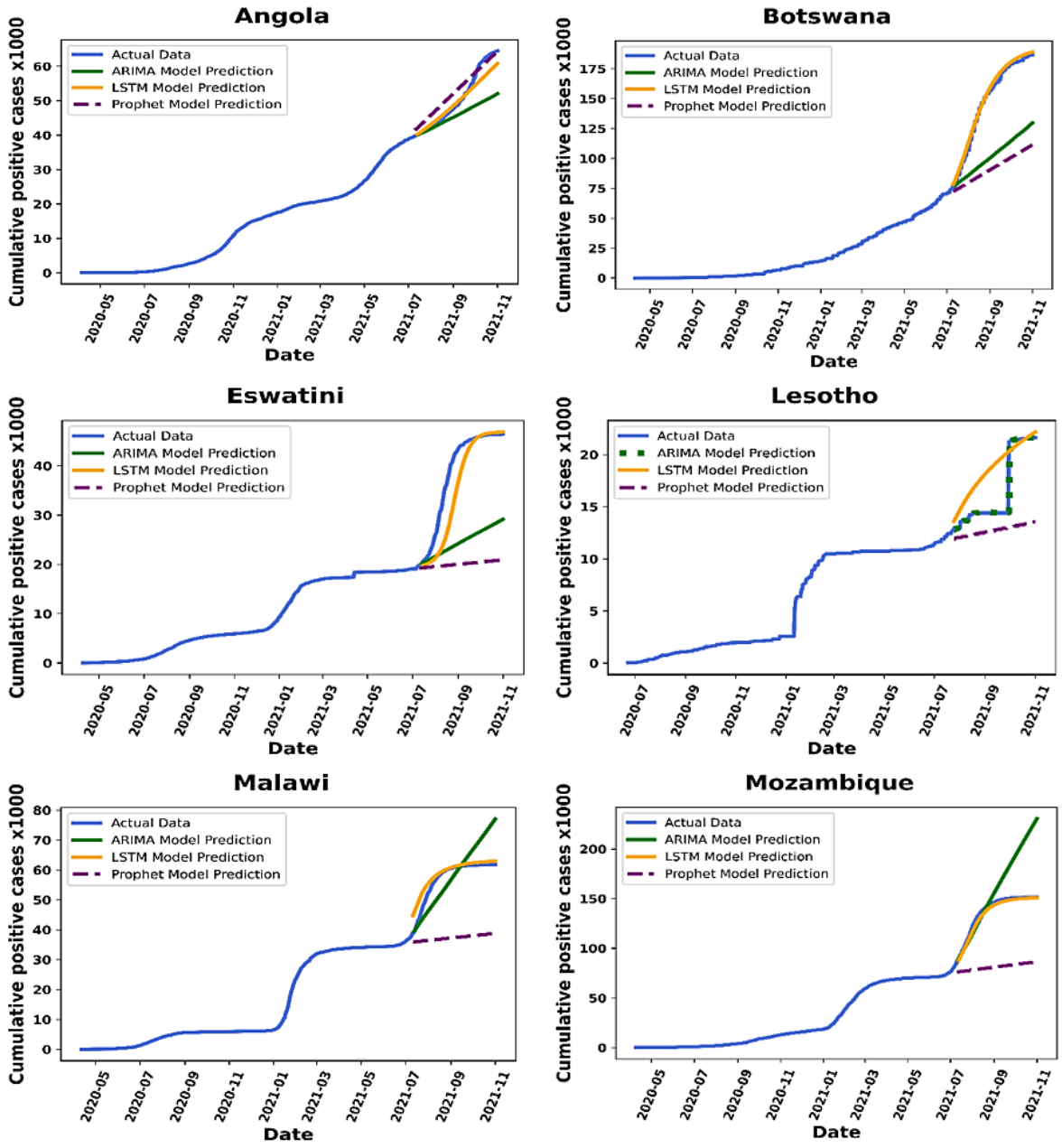

Figure 4.7. Actual and predicted cumulative positive cases in Southern Africa(a)

The LSTM algorithm generated the best-matched estimation results for various nations (Botswana, Malawi, and Mozambique) as shown in Figure 4.7. The ARIMA model outperformed the other two techniques in Lesotho. The Prophet model performed poorly in four nations: Malawi, Mozambique, Eswatini, and Lesotho. This algorithm forecasts a generally fixed number of instances in these nations, with modest rises in the expected number of instances.

In Angola, the LSTM and Prophet models provided slightly similar forecasts that were close to the actual data. However, the ARIMA algorithm projected a smaller number of incidents, which was very distinct but somewhat similar to the true data. In this nation, the three algorithms' expected outcomes are quite consistent. This can be linked to Angola's steady increase in the number of COVID-19 instances, making it simpler for all algorithms to extract the inherent data associations and patterns in order to produce better forecasts.

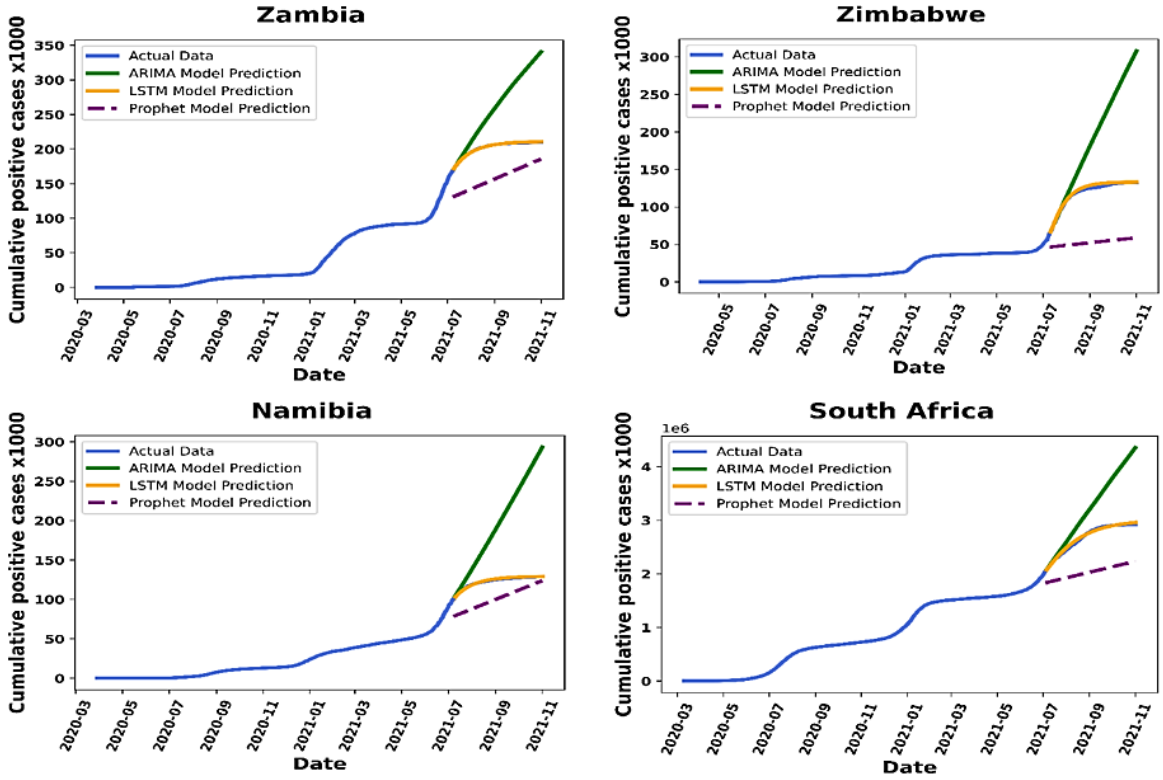

Figure 4.8. Actual and predicted cumulative positive cases for Southern Africa(b)

While comparing the forecasting models to the other nations, Figure 4.8 shows that they did the worst. This algorithm generated projections that were significantly less reliable when compared to the real data. The ARIMA algorithm projected more COVID-19 incidences in all four nations than any of the other algorithms.

The LSTM algorithm also appears to have superior results in terms of fitting projections. The Prophet algorithm, which has the second-best predictive accuracy, comes in second to the LSTM algorithm. As seen in both figures 4.7 and 4.8, the LSTM model gives the perfect general predictive performance in the South African region, whereas the ARIMA algorithm provides the lowest forecast outcomes.

Table 4.3. Model performance in Southern Africa

Country	Model	PSNR	R value	NRMSE	SMAPE	RMSE	MSE	MAPE
South-Africa	ARIMA	-68.6754	-5.7832	0.7905	16.5715	692322.2002	479310000000.0000	19.0097
	LSTM	-39.7648	0.9913	0.0283	0.7800	24818.6368	615964732.6000	0.7830
	Prophet	-68.2322	-5.1251	0.7512	26.7633	657882.1667	432809000000.0000	23.5047
Zambia	ARIMA	-48.9441	-48.6151	1.7532	23.3656	71407.6662	5099054792.0000	28.1140
	LSTM	-6.7122	0.9970	0.0136	0.2252	552.2662	304997.9557	0.2252
	Prophet	-44.8702	-18.4190	1.0968	24.6496	44673.6041	1995730903.0000	21.8100
Namibia	ARIMA	-50.7674	-168.8737	3.0293	40.7690	88086.5415	7759238793.0000	57.0550
	LSTM	-11.0236	0.9820	0.0312	0.6692	907.2334	823072.4421	0.6697
	Prophet	-39.0956	-10.5598	0.7902	19.9281	22978.5282	528012758.2000	17.8522
Eswatini	ARIMA	-35.7542	-1.6589	0.5871	40.7329	15640.5055	244625412.3000	32.6660
	LSTM	-25.7619	0.7336	0.1858	10.6050	4950.3420	24505885.9200	9.5222
	Prophet	-38.0908	-3.5536	0.7683	58.2823	20468.2129	418947739.3000	43.3898
Lesotho	ARIMA	-8.5919	0.9618	0.0775	0.6781	685.6982	470182.0215	0.6157
	LSTM	-21.4589	0.2614	0.3409	14.5044	3016.3651	9098458.4170	16.4016
	Prophet	-25.6834	-0.9537	0.5545	23.6798	4905.8203	24067072.8200	20.2240
Malawi	ARIMA	-28.4637	-0.0858	0.2957	9.3612	6756.5296	45650692.2400	9.5218
	LSTM	-20.0071	0.8451	0.1117	3.3325	2552.0723	6513073.0240	3.4858
	Prophet	-38.2399	-9.3125	0.9113	41.4722	20822.5839	433580000.3000	34.0048
Mozambi-que	ARIMA	-42.8290	-2.4604	0.5430	14.7160	35317.3358	1247314208.0000	16.9596
	LSTM	-21.2013	0.9762	0.0450	1.9387	2928.2214	8574480.5670	1.9128
	Prophet	-47.2140	-8.4979	0.8996	50.2045	58511.4488	3423589641.0000	39.7238
Botswana	ARIMA	-45.9388	-1.1362	0.4543	34.8145	50522.0752	2552480083.0000	29.1172
	LSTM	-21.9147	0.9915	0.0286	1.9982	3178.8901	10105342.2700	2.0284
	Prophet	-47.6825	-2.1915	0.5553	44.8292	61753.7964	3813531370.0000	35.9618
Angola	ARIMA	-28.5294	0.3054	0.2763	9.3038	6807.8362	46346633.7300	8.6089
	LSTM	-19.7646	0.9077	0.1007	3.0467	2481.8282	6159471.2140	2.9731
	Prophet	-21.5315	0.8613	0.1234	5.5594	3041.7034	9251959.5740	5.7561
Zimbabwe	ARIMA	-50.9603	-25.8469	1.3262	38.1986	90064.6700	8111644782.0000	54.1828
	LSTM	-19.8316	0.9793	0.0368	1.6663	2501.0312	6255157.0630	1.6868
	Prophet	-48.4614	-14.1010	0.9947	75.8921	67547.6906	4562690505.0000	54.7687

The evaluation criteria utilized to establish the best forecasting models in the Southern African region are shown in Table 4.3. The LSTM algorithm obtained the largest R value of 0.9970 in Zambia, whereas the ARIMA algorithm produced the lowest R value of -168.8737 in Namibia. The ARIMA algorithm, on the contrary, recorded the lowest PSNR value of -68.6754 in South Africa, whereas the LSTM algorithm produced the largest PSNR value of -6.7122 in Zambia. The LSTM algorithm turned out to be the best algorithm that predicts in this region according to the PSNR and R criteria, whereas ARIMA shows the worst performance.

The RMSE metric ranges of 552.2662 - 24818.6368, 685.6982 - 692322.2002 and 3041.7034 - 657882.1667 were obtained by the LSTM, ARIMA and LSTM models, respectively. It is also evident from this metric that the best RMSE range was produced by the LSTM model compared to the rest of the models. With the smallest value of the MAPE metric of 0.2252 from Zambia, the overall best performing model in the Southern Africa region is observed to be the LSTM model, while the worst performing model, with the largest MAPE of 57.0550 from Namibia, is observed to be the ARIMA model.

4.1.4. Western Africa

Twelve nations from this region were utilized in this study. Nigeria has the highest number of inhabitants in this region, with 206139589 individuals. Guinea-Bissau, on the contrary, does have the lowest population, with a population of 198,001 inhabitants.

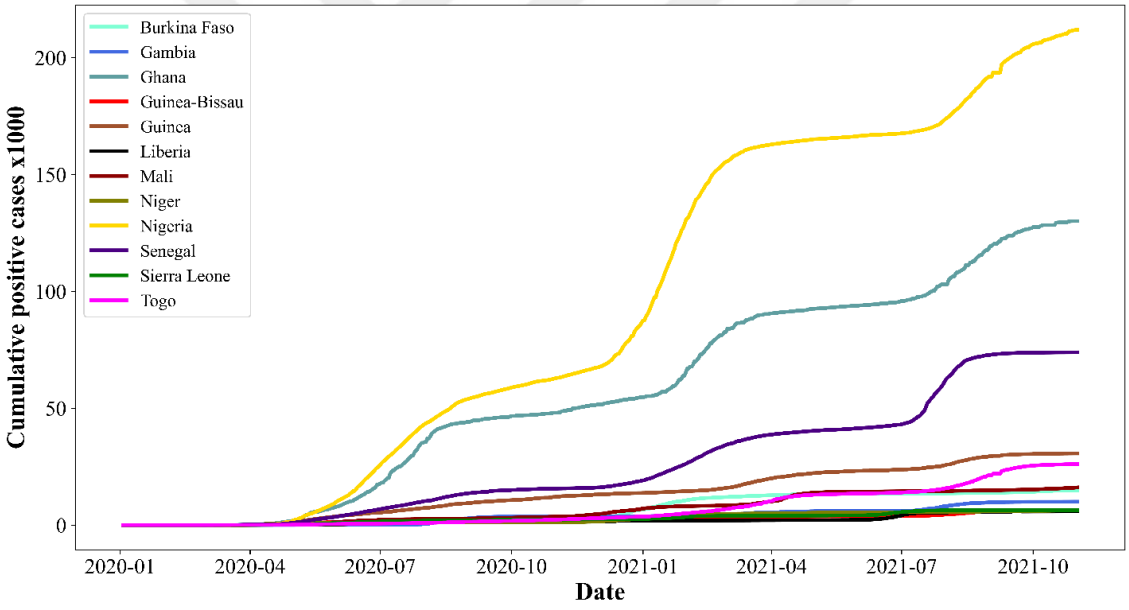

Figure 4.9. Cumulative positive cases in the Western African region

A comparison graph of the 12 nations included in this research from this region, is shown in Figure 4.9. The condition of the COVID-19 epidemic in each of the 12 counties is depicted here. It also shows the intensity of the risk condition in terms of COVID-19 propagation. There were almost no COVID-19 instances recorded in this region throughout January and April of the same year. Nonetheless, after April of that year, the very first COVID-19 occurrences started to be documented.

Following this, there is a substantial spike in the number of incidents in around 4 nations, including Nigeria, Ghana, Senegal, and Mali, whereas the number of incidents in the

remaining nations gradually increases. Nigeria had the greatest number of instances throughout time, preceded by Ghana and Senegal. Nigeria, with a population of more than 200 million and the largest number of COVID-19 incidents, is the region's most risky nation. If prompt action is not taken, there is a greater risk of the virus spreading to other nations.

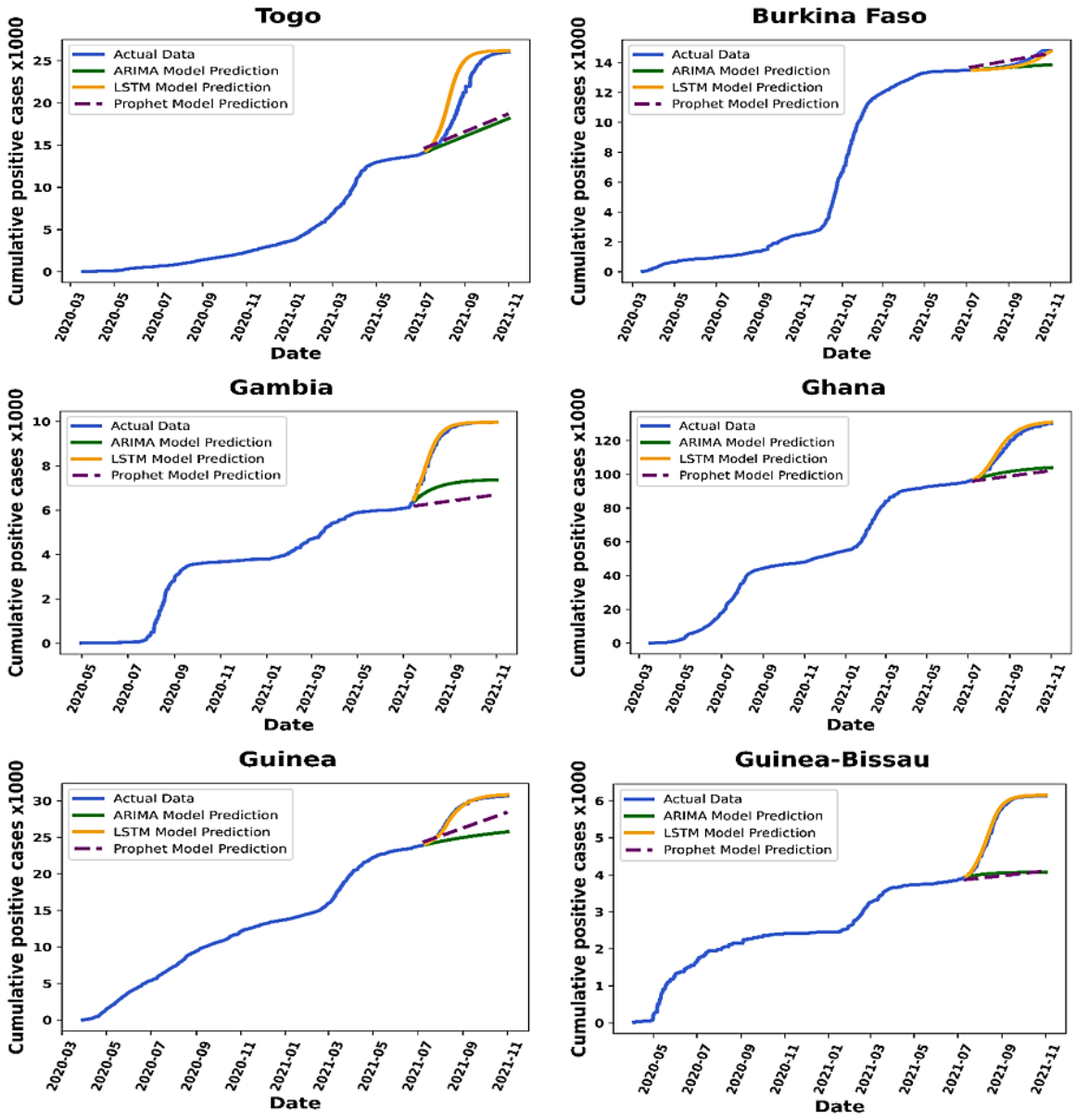

Figure 4.10. Actual and predicted cumulative positive cases for Western Africa(a)

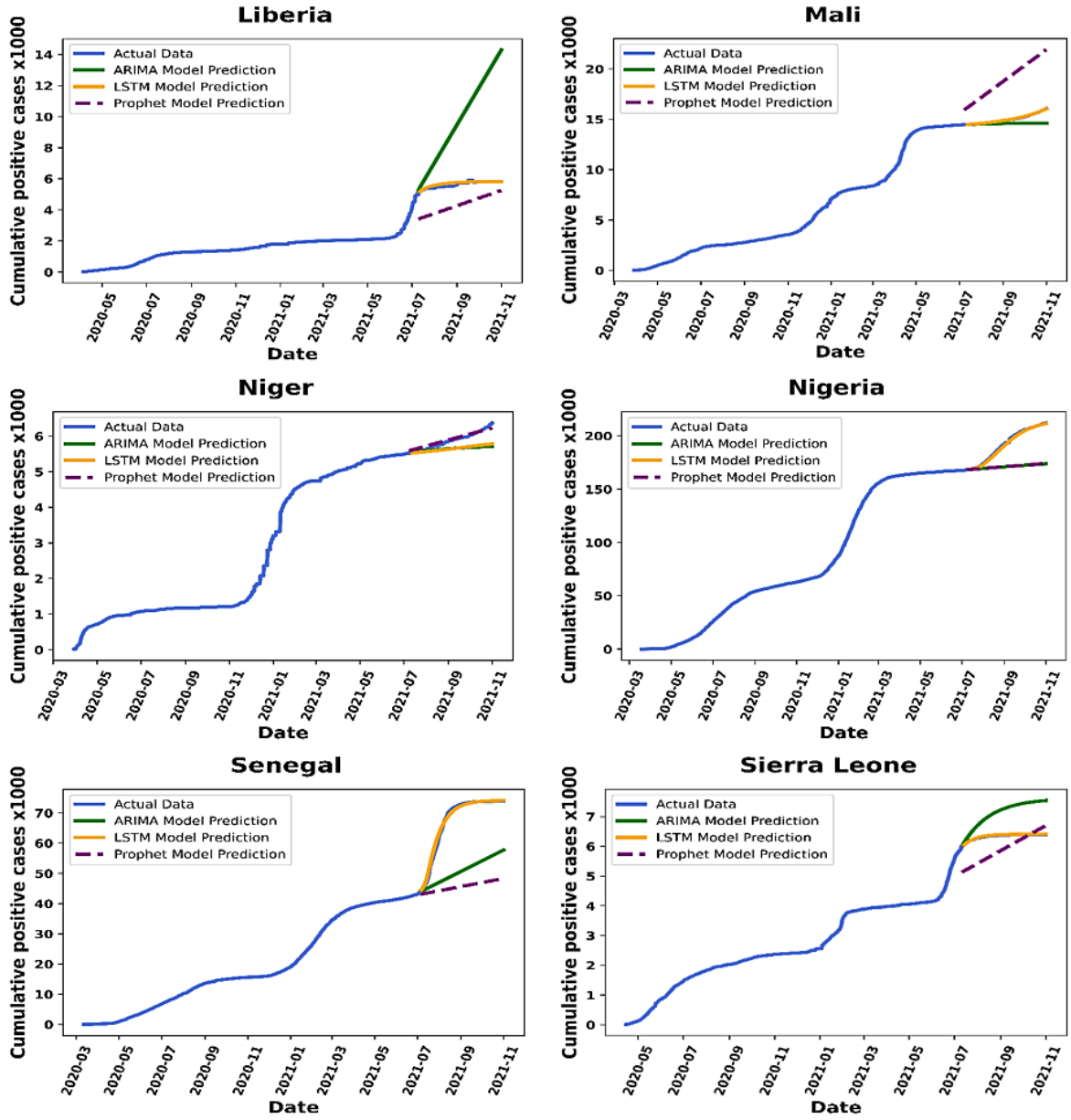

Figure 4.11. Actual and predicted cumulative positive cases in Western African(b)

The outcomes of the three models' predictions in the Western African region are shown in Figure 4.10. The LSTM algorithm surpassed the other two algorithms in providing the best possible perfectly matched predictive accuracy in the very first batch of nations from this region. This is evident in nations such as Guinea, Guinea-Bissau, Gambia, Ghana, and Togo. The Prophet model makes the most accurate predictions in Burkina Faso. In three countries: Guinea-Bissau, Ghana, and Togo, the ARIMA and Prophet have been witnessed making matching forecasts. When compared to the actual data, these projections show a smaller COVID-19 number of cases. The second set of model projections for the Western African region is shown in figure 4.11. In Niger, the Prophet approach has the highest accuracy in terms of prediction performance. If contrasted to the other nations, this model

performed better. It could also be deduced from this graph that the ARIMA algorithm did not perform well in any of the nations. The LSTM algorithm continues to be the top scoring approach in this region, with the best-matched model performance in all six nations. Both the ARIMA and Prophet approaches in Nigeria produce similar forecasts, although they are still lower and notably distinct from the real data. These findings show that the LSTM model is an excellent model for predicting in the West African region.

Table 4.4. Performance of the models in Western Africa

Country	Model	PSNR	R value	NRMSE	SMAPE	RMSE	MSE	MAPE
	ARIMA	-1.1740	-0.6132	0.3475	3.9254	291.9044	85208.1787	3.8076
Niger	LSTM	-0.5325	-0.3917	0.3228	3.8508	271.1221	73507.1931	3.7488
	Prophet	12.2196	0.9262	0.0743	0.9621	62.4539	3900.4896	0.9659
	ARIMA	-7.2888	-0.7804	0.3645	2.8408	590.1751	348306.6487	2.7670
Mali	LSTM	19.7407	0.9965	0.0162	0.1374	26.2729	690.2653	0.1375
	Prophet	-24.2073	-86.5715	2.5566	22.7287	4139.1138	17132263.0500	25.9383
	ARIMA	-25.5340	-558.9784	5.2358	49.7815	4822.1454	23253086.2600	72.6324
Liberia	LSTM	7.4258	0.7167	0.1178	1.4477	108.4557	11762.6389	1.4647
	Prophet	-14.4240	-42.3685	1.4571	26.6279	1341.9666	1800874.3560	23.2167
	ARIMA	-23.7940	-1.8990	0.5884	12.7953	3946.7398	15576755.0500	11.8515
Guinea	LSTM	3.7225	0.9949	0.0248	0.5042	166.1170	27594.8577	0.5034
	Prophet	-19.6152	-0.1076	0.3637	7.5991	2439.4969	5951145.1250	7.2484
	ARIMA	-15.9092	-2.9237	0.7195	28.2155	1592.2175	2535156.5670	23.9589
Guinea-Bissau	LSTM	13.1228	0.9951	0.0254	0.8095	56.2861	3168.1251	0.8156
	Prophet	-16.0789	-3.0801	0.7337	29.4110	1623.6256	2636160.0890	24.9430
	ARIMA	-36.9592	-1.3960	0.5336	13.5331	17967.9858	322848513.7000	12.4038
Ghana	LSTM	-18.7581	0.9637	0.0656	1.5789	2210.2812	4885342.9830	1.5967
	Prophet	-37.8953	-1.9724	0.5943	15.6611	20012.7714	400511019.1000	14.2355
	ARIMA	-18.8193	-3.6199	0.6265	24.9010	2225.9044	4954650.3980	21.8765
Gambia	LSTM	9.2716	0.9928	0.0247	0.7308	87.6925	7689.9746	0.7358
	Prophet	-21.1720	-6.9414	0.8214	34.6748	2918.3552	8516797.0730	29.1977
	ARIMA	-4.7163	-0.0108	0.3379	2.0304	438.8913	192625.5732	1.9840
Burkina Faso	LSTM	3.2095	0.8370	0.1357	0.9683	176.2237	31054.7924	0.9607
	Prophet	-0.2520	0.6384	0.2021	1.7649	262.5062	68909.5050	1.7794
	ARIMA	-27.1499	-0.6747	0.4841	23.3990	5808.0929	33733943.1400	20.0786
Togo	LSTM	-18.9326	0.7475	0.1880	8.1182	2255.1162	5085549.0760	8.7542
	Prophet	-26.4842	-0.4367	0.4484	21.2872	5379.6047	28940146.7300	18.5120
	ARIMA	-10.4028	-88.3031	2.1384	11.3549	844.6563	713444.2651	12.1692
Sierra Leone	LSTM	22.5264	0.9545	0.0483	0.2727	19.0642	363.4437	0.2728
	Prophet	-6.9960	-39.7560	1.4446	8.0593	570.6140	325600.3370	7.6311
	ARIMA	-39.5807	-1.5257	0.5510	10.7816	24298.1514	590400161.5000	10.0128
Nigeria	LSTM	-12.3440	0.9952	0.0239	0.4389	1056.1850	1115526.7540	0.4374
	Prophet	-39.5353	-1.4995	0.5481	10.7166	24171.6715	584269703.1000	9.9560
	ARIMA	-36.9374	-2.5463	0.5945	27.3320	17923.0581	321236011.7000	23.6709
Senegal	LSTM	-9.9932	0.9928	0.0267	0.9606	805.7498	649232.7402	0.9669
	Prophet	-39.2209	-4.9995	0.7732	37.3071	23312.2755	543462189.0000	30.8797

Table 4.4 shows the forecast findings for the 12 Western African nation using the seven indicators utilized in this research. The LSTM algorithm from Sierra Leone, with a PSNR score of 22.5264, had the finest predictive accuracy, as seen in the table.

Nigeria, on the contrary, had the worst score, with a PSNR score of -39.5807, according to the ARIMA algorithm. The R measure also shows that the LSTM algorithm in Mali yielded the greatest score of 0.9965, while the ARIMA algorithm in Liberia produced the minimum score of -558.9784.

It's important to note that greater PSNR and R values indicate improved results. In the Western African region, the LSTM algorithm performed better than the other two algorithms. The smallest RMSE score of 19.0642 produced by the same algorithm in Sierra Leone supports this fact, whereas the ARIMA algorithm in Nigeria has the poorest performance with this statistic with a score of 24,298.1514.

If the MAPE statistic values from across all three algorithms are compared, the LSTM model delivers the best range of 0.1375–8.7542, preceded by the ARIMA algorithm with a range of 1.9840–72.6324 and the Prophet algorithm with a range of 0.9659–30.8797. The LSTM algorithm has an outstanding outcome in the Western region, according to these performance metrics.

4.1.5. Eastern Africa

Twelve nations were investigated within this region. The Comoros have the smallest population density, with 869601, whereas Ethiopia has the largest population density, with 114963588 people at the time of this research.

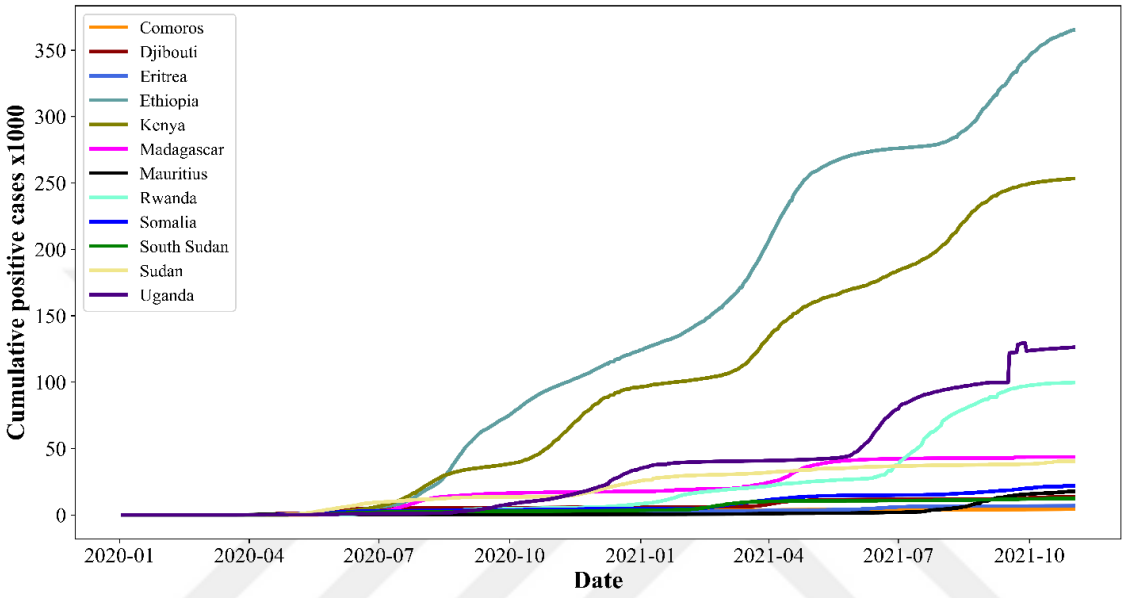

Figure 4.12. Cumulative positive cases in the Eastern African region

The chart in figure 4.12, shows the cumulative positive COVID-19 instances for the nations in Africa's Eastern region. Ethiopia has the largest number of COVID- 19 instances in this region, preceded by Kenya. It's noteworthy that Kenya's population density, at 53771296 inhabitants, closely preceeds Ethiopia's, whereas its number of COVID-19 cases also closely matches Ethiopia's. This further implys that there is an approximately significant correlation between the population density and the COVID-19 case number.

If preemptive actions are not taken, the Eastern region faces a greater chance of COVID-19 spreading at a faster rate. The incidence of the first cases in the region occurred fairly late, as evidenced by the fact that large number of reported cases began to be documented in all nations soon just after month of July in 2020. Kenya has the largest number of COVID-19 transmission episodes in this region.

With the exception of Ethiopia, Kenya, Uganda, Rwanda, Madagascar, and Sudan, the remainder of the nations have seen a modest surge in the prevalence of the COVID-19 cases.This could be due to a variety of policies implemented by the different nations as well as the public at large. For instance, the Comoros, the region's least populous nation.

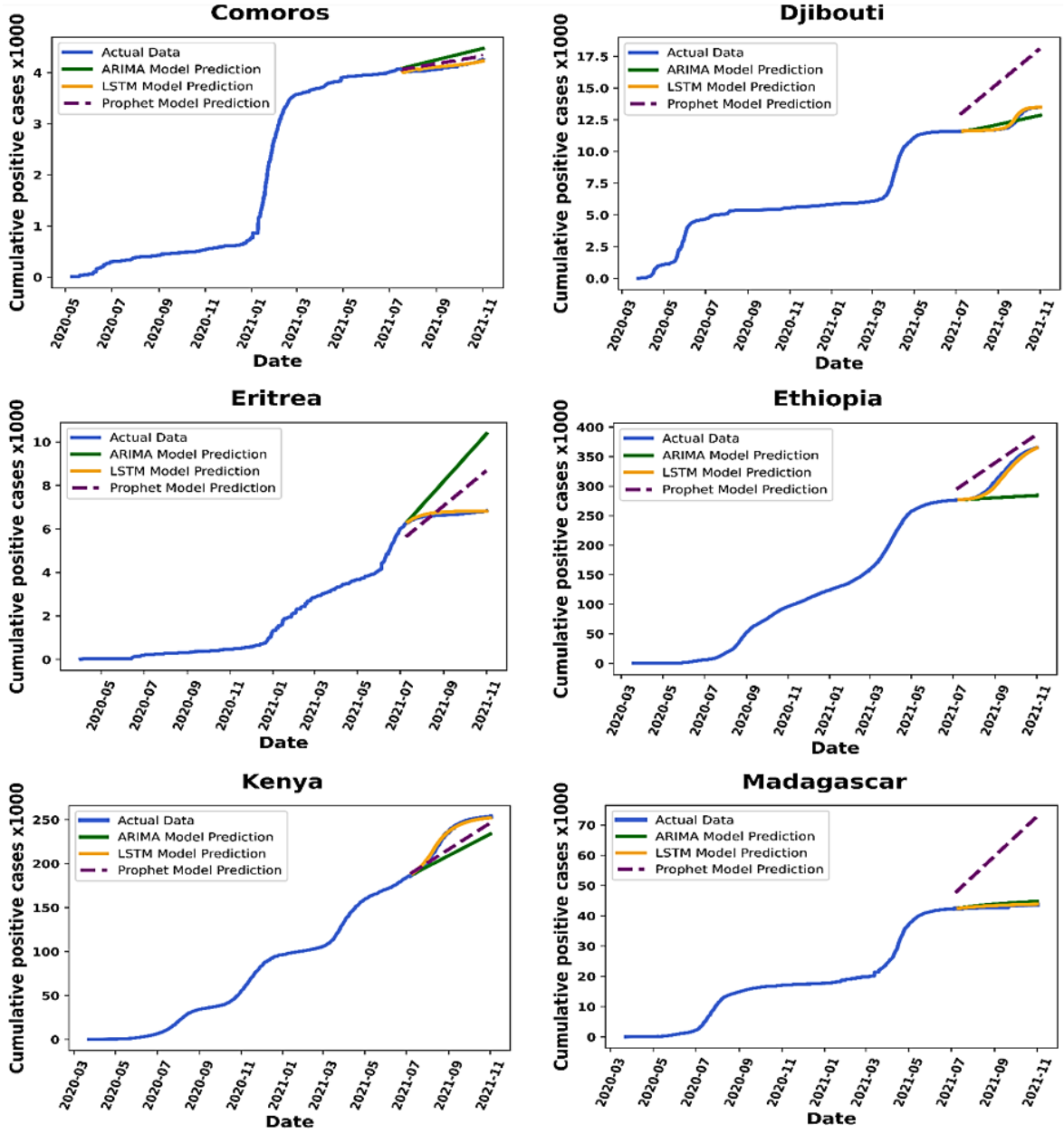

Figure 4.13. Actual and predicted cumulative positive cases in Eastern Africa (a)

As seen in Figure 4.13, the Prophet approach has the weakest predictive accuracy in both Djibouti and Madagascar. The LSTM algorithm, on the contrary, clearly achieves the highest predictive accuracy in all nations involved, as shown by the same graph.

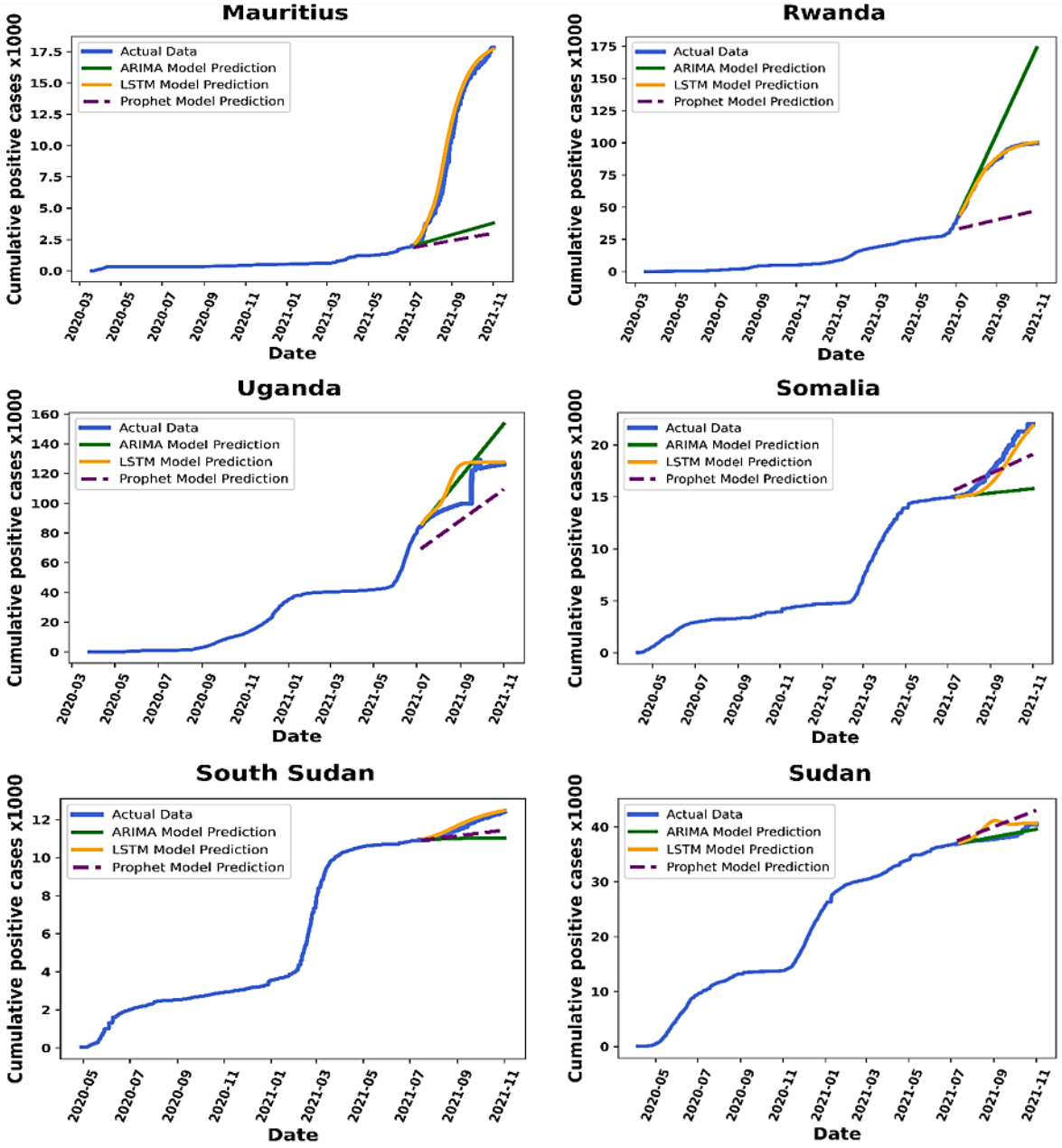

Figure 4.14. Actual and predicted cumulative positive cases in Eastern Africa(b)

Figures 4.13 and 4.14 show the LSTM, ARIMA, and Prophet algorithms' prediction outcomes in the 12 Eastern African nations investigated in this research. Both the projected data from the algorithms and the expected real data are plotted in these outcomes. Figure 4.13 shows that in the Comoros, all three algorithms did reasonably well, preceded by Sudan, as seen in figure 4.14. The three algorithms reveal considerable relative disparities in performance in the remaining nations, as seen in both graphs. When contrasted to the Prophet method in Madagascar, both the LSTM and ARIMA algorithms produced improved matching forecasting outcomes. When contrasted to the ARIMA and Prophet algorithms, the LSTM algorithm has the best possible corresponding predictive accuracy in figure 4.14.

Both the ARIMA and Prophet algorithms also have the worst predictive accuracy in Mauritius and Rwanda. In the eastern area, the LSTM model outscored the ARIMA and Prophet algorithms.

Table 4.5. Performance parameters of the models for Eastern Africa

Country	Model	PSNR	R value	NRMSE	SMAPE	RMSE	MSE	MAPE
	ARIMA	-30.9283	-1.5722	0.5692	93.1564	8973.3930	80521781.9300	59.9527
Mauritius	LSTM	-9.8988	0.9797	0.0506	7.7913	797.0358	635266.0665	8.3734
	Prophet	-31.3969	-1.8653	0.6007	104.3478	9470.7650	89695389.6900	65.2334
	ARIMA	-10.3860	-2.8964	0.6832	1.7484	843.0209	710684.2378	1.7672
Madagascar	LSTM	-2.1830	0.4107	0.2657	0.6019	327.8600	107492.1796	0.6048
	Prophet	-37.2296	-1882.7825	15.0212	32.4367	18536.1595	343589209.0000	39.7624
	ARIMA	-38.5762	0.1094	0.3237	8.2885	21644.5874	468488163.7000	7.8676
Kenya	LSTM	-16.2976	0.9947	0.0249	0.6621	1665.0265	2772313.2460	0.6640
	Prophet	-34.9861	0.6104	0.2141	5.2961	14316.8267	204971526.8000	5.1147
	ARIMA	-42.6049	-3.1725	0.6113	22.6311	34418.0092	1184599357.0000	27.5678
Rwanda	LSTM	-11.4368	0.9968	0.0169	0.9610	951.4349	905228.3689	0.9569
	Prophet	-44.8293	-5.9637	0.7897	66.9946	44463.6562	1977016723.0000	49.7882
	ARIMA	-17.9013	-234.2806	3.3829	21.4673	2002.6522	4010615.8340	25.1679
Eritrea	LSTM	8.3570	0.4431	0.1646	1.2999	97.4299	9492.5854	1.3106
	Prophet	-11.2635	-50.0280	1.5754	10.6952	932.6455	869827.6287	11.4391
	ARIMA	-45.0386	-1.0228	0.5137	11.0389	45548.2859	2074646348.0000	10.0965
Ethiopia	LSTM	-25.4272	0.9779	0.0537	1.1764	4763.2242	22688304.7800	1.1653
	Prophet	-40.5723	0.2767	0.3072	8.1788	27236.8169	741844194.8000	8.5723
	ARIMA	2.7683	-6.6139	0.7476	4.2035	185.4063	34375.4961	4.3021
Comoros	LSTM	18.6599	0.8039	0.1200	0.5967	29.7540	885.3005	0.5992
	Prophet	8.3879	-1.0876	0.3915	2.2780	97.0829	9425.0895	2.3056
	ARIMA	-3.8103	0.6871	0.2116	2.6073	395.4184	156355.7111	2.5933
Djibouti	LSTM	7.8871	0.9788	0.0550	0.4601	102.8451	10577.1146	0.4634
	Prophet	-22.6304	-22.8474	1.8469	23.6379	3451.9329	11915840.7500	27.0253
	ARIMA	-35.2449	0.0511	0.3294	10.4174	14749.7766	217555909.7000	11.2426
Uganda	LSTM	-34.4800	0.2044	0.3017	8.4224	13506.4698	182424726.5000	9.2481
	Prophet	-37.2039	-0.4897	0.4128	17.6531	18481.4583	341564300.9000	16.1231
	ARIMA	-9.3188	-1.3764	0.4994	5.1883	745.5587	555857.7751	4.9893
South-Sudan	LSTM	3.9456	0.8879	0.1084	1.2819	161.9053	26213.3262	1.2915
	Prophet	-6.4872	-0.2381	0.3604	3.6874	538.1478	289603.0546	3.5850
	ARIMA	-22.5747	-1.0788	0.4898	15.0670	3429.8690	11764001.3600	13.4804
Somalia	LSTM	-10.2801	0.8774	0.1189	3.9871	832.8056	693565.1674	3.8825
	Prophet	-15.7176	0.5714	0.2224	6.7348	1557.4683	2425707.5060	6.5082
	ARIMA	-4.6451	0.8122	0.1243	0.8412	435.3047	189490.1818	0.8396
Sudan	LSTM	-17.3794	-2.5243	0.5385	3.9144	1885.8611	3556472.0890	4.0335
	Prophet	-18.5771	-3.6434	0.6181	5.0666	2164.6931	4685896.2170	5.2195

The 3 different model performance outcomes for the 12 nations from Africa's eastern region are listed in table 4.5. The algorithms' predictive accuracy was measured using seven criteria. PSNR and R value are the measurements that perform well for larger values. Besides

this, the remainder of the criteria show the best results for smaller figures. The LSTM algorithm achieved the largest score of 18.6599 from the Comoros, according to the PSNR statistic. Using the same measures, it is clear that the ARIMA model had the worst outcome in Ethiopia, with the lowest PSNR value of -45.0386. This indicates that the ARIMA model's forecasted data contains considerably more noise, which directly translates to poor performance. This is also shown in figure 4.14, where the ARIMA projections deviate further from the real data. The Prophet, with an R value of -1882.7825, has the lowest R value in Madagascar.

When contrasted with different model predictions and their corresponding actual numbers in both Madagascar and the other nations, the Prophet model's predicted values in Madagascar show the weakest connection with the real data in this nation. Using the same criteria, the LSTM model in Rwanda was found to have the best and largest value of 0.9968. As seen in figure 4.14 in Rwanda, there is a high correlation between the LSTM projected instances and the real cases. When the RMSE values are sorted, the LSTM algorithm has the best outcome of 29.7540–13506.4698, preceded by the Prophet algorithm with values ranging from 97.0829–44463.6562 and the ARIMA algorithm with a range of 185.4063–45548.2859. In this region, it is evident that the LSTM algorithm is the most reliable predictive model.

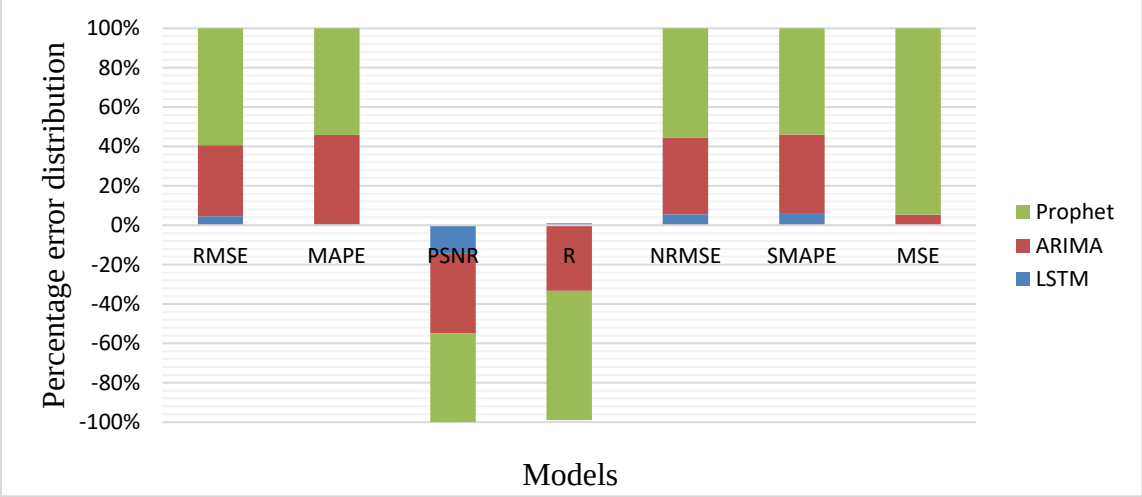

Figure 4.15. Total error distribution of the models

The respective model predictive performance of regions is shown in figure 4.15. It displays percentage proportions, including both positive and negative orientations, to measure every performance of the models in terms of how it contributes to the overall error score for the 7 performance criteria examined. Across both PSNR and R, excellent performance is

represented by a much more positive percentage proportion, while poor performance is represented by a more negative percentage proportion. The lower percentage proportions inclined in the positive orientation indicate satisfactory accuracy for RMSE, MAPE, NRMSE, SMAPE, and MSE errors.

A significant positive proportion, on the contrary, indicates poor predictive accuracy. The LSTM model had the greatest overall results in this investigation, preceded by the ARIMA model, and finally the Prophet model. It is because, in all 5 criteria, the LSTM algorithm has the least proportion of total error. The ARIMA algorithm comes next, with percentage proportions that are greater than the LSTM algorithm but less than the Prophet algorithm. The PSNR and R scores also show that the LSTM algorithm significantly outperformed the rest of the algorithms.

The LSTM model's PSNR and R scores are all positive, indicating that it outperformed the ARIMA and Prophet algorithms in these 2 criteria. It is then preceded by the ARIMA algorithm and, finally, the Prophet algorithm. The effectiveness of the LSTM model is due to its ability to analyze and manage all types of sequential data, whereas the other two algorithms are limited by the reliability of the underlying intrinsic data attributes. The ARIMA algorithm fits better with stationary data and demands a greater quantity of data, which also makes it perform poorly when data is not relatively reliable. Because the COVID-19 outbreak is indeed a new problem with few data available, the data employed in this investigation was limited.

Notwithstanding differencing efforts throughout the ARIMA model fitting, the datasets in most nations were not substantially able to be made stable. If contrasted to the LSTM algorithm, all of these characteristics lead to its poor results. The Prophet algorithm, on the contrary, is found to be the generally poorest performing model in this research. Due to the minimal data collected and provided for training, this Fourier series-based algorithm failed to discover and understand important trends, seasonality, and holiday patterns within the data to generate best-matching projections, considering its simplicity of implementation and lack of data preparation requirements. Because the LSTM model has numerous hyperparameter tuning features, it can be fine-tuned until the optimum outcome is achieved. If contrasted with other different algorithms, the LSTM model had the most computation and time sophistication in order to attain optimum performance.

4.2. Forecasting for the next 61 days

The second significant component of this research included the forecasting of the cumulative positive cases by the best algorithm for every nation for a duration of 61 days upon identifying the best model. The last date of confirmed incidents for each nation in all areas was 2021-11-1 at the time of access to the main data source employed in this research. For every nation in the regions of the African continent, the COVID-19 cases were forecasted from the last date in the source dataset through to the date of 2022-01-02.

4.2.1. Northern Africa

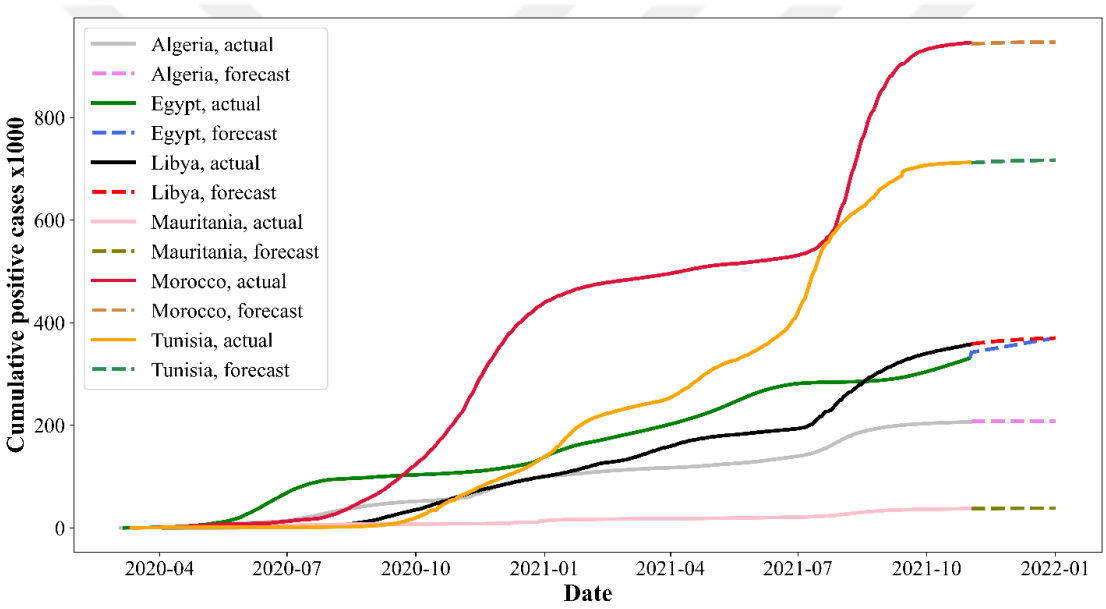

Figure 4.16. Observed and forecasted COVID-19 cases for Northern Africa

As shown in Figure 4.16, the COVID-19 cases are predicted to increase at a rapid rate in Egypt. Whereas COVID-19 incidents are projected to keep a leveled rate of increase in nations like Tunisia, Algeria, and Mauritania, they are projected to continue rising in Libya. A noticeable, modest drop is projected in Morocco, preceded by a stable number of COVID-19 instances with a minor rise at the end of the projected timeframe.

Every one of the Northern African nations that recorded COVID-19 cases is predicted to have a rise by the conclusion of the forecast timeframe. COVID-19 incidences are anticipated to rise from 206452 to 208009, 37320 to 38250, 712747 to 716835, 331017 to 370164, 357338 to 369986, and 946145 to 947226 in Algeria, Mauritania, Tunisia, Egypt, Libya, and Morocco, correspondingly. Egypt is the nation in this region with the highest

predicted rise in COVID-19 cases, with a rise of 11.83 percent of the total number of COVID-19 instances at the conclusion of the forecast timeline.

4.2.2. Central Africa

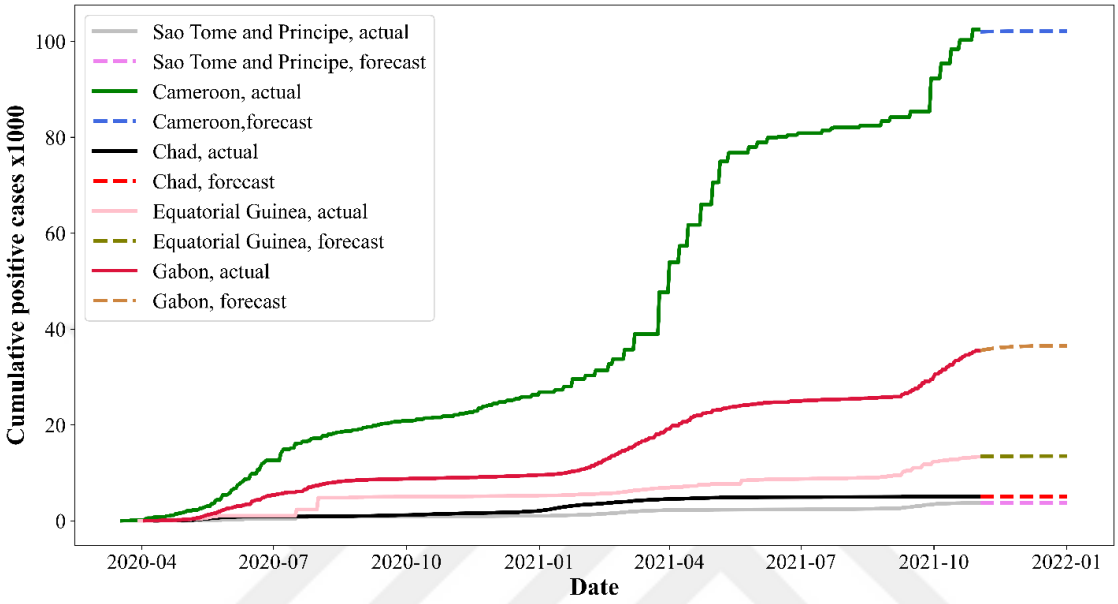

Figure 4.17. Observed and forecasted COVID-19 cases for Central Africa.

The forecasted COVID-19 incidents for the five Central African nations are depicted in figure 4.17. In Cameroon, the rate of spread is likely to slow modestly before returning to its previous level. A modest increase is projected in Gabon and Equatorial Guinea, whereas a consistent rate of spread in the COVID-19 incidences is anticipated in Chad and So Tomé and Principe. In Cameroon, the number of COVID-19 incidents is predicted to drop from 102499 to 102129 by the end of the forecasted period. Cameroon is the only country in Central Africa where the case number is predicted to decline.

The remainder of the nation is also projected to see an increase in the number of cases. In Gabon, Chad, Equatorial Guinea, and São Tomé and Príncipe, COVID-19 incidences are predicted to rise from 35525 to 36522, 5069 to 5072, 13368 to 13508 and 3714 to 3717, correspondingly. Gabon is predicted to see the biggest increase in case numbers in this region, with a percentage rise of 2.81 percent.

4.2.3. Southern Africa

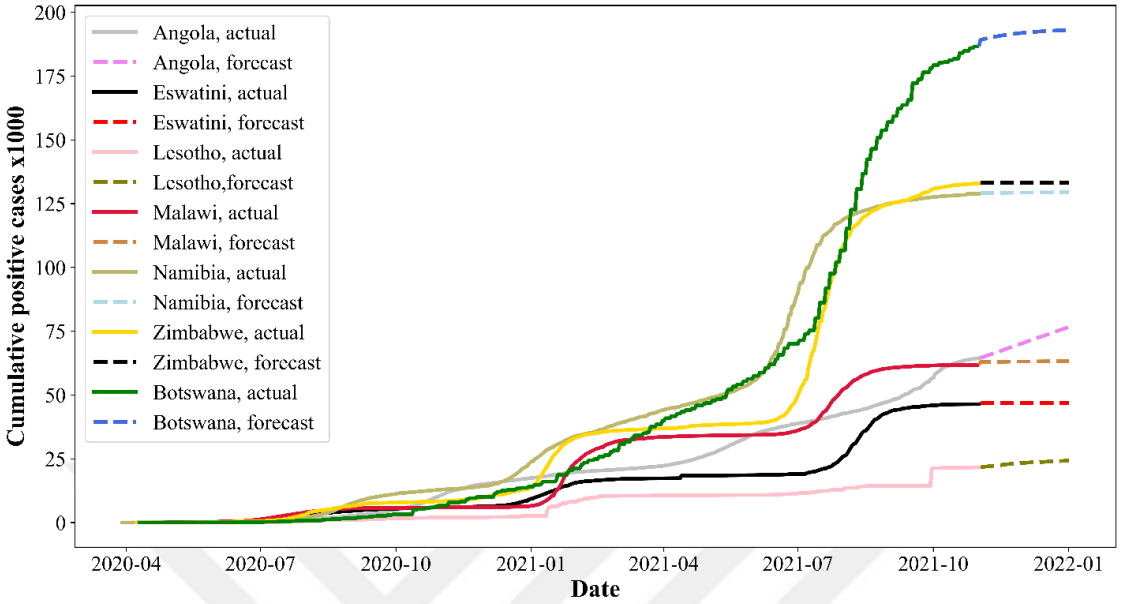

Figure 4.18. Observed and forecasted COVID-19 cases for Southern Africa(a)

Nations in the Southern African region were divided into two graphs to display the anticipated cumulative cases for comprehension. This is due to the fact that the number of incidents in South Africa is significantly higher than in other countries in the same area. As a result, graphs for other countries would be jumbled together and impossible to evaluate.

A depiction of the observed and anticipated total cases for seven Southern African nations is shown in Figure 4.18. According to this graph, the predicted rate of increase in the COVID-19 cases in Angola is bigger than in the other nations. Following Angola, Lesotho has seen a substantial growth in COVID-19 incidents. Botswana follows Lesotho, with a minor but noticeable rise in the number of cases. Besides these three, the remaining nations appear to establish a steady number of cases with minor fluctuations.

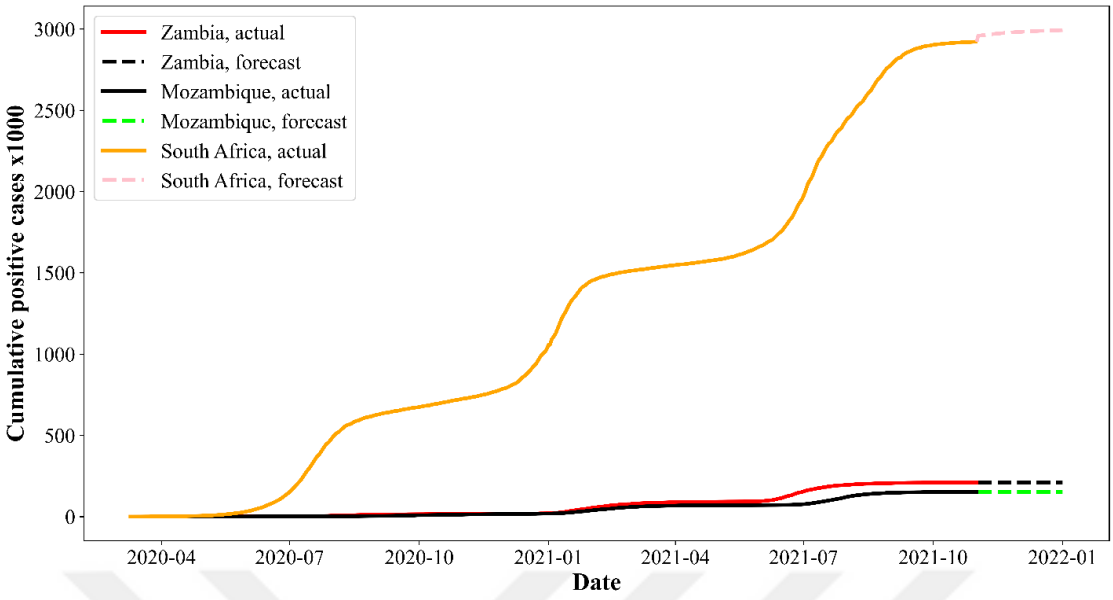

Figure 4.19. Observed and forecasted COVID-19 cases for Southern Africa(b)

Figure 4.18 is an extension of Figure 4.19, which also depicts a plot of projected and observed instances for 3 Southern African nations. The number of cumulative cases is projected to remain stable in both Zambia and Mozambique, whereas a considerable progressive rise is anticipated in both countries. Only Mozambique's COVID-19 cumulative cases are predicted to decline from 151292 to 151051 at the conclusion of the forecast timeline among the nations of this region. The number of cases is projected to rise in the remaining countries.

The number of COVID-19 incidents is predicted to rise from 64433 to 76655, 186594 to 193024, 61796 to 63201, 128886 to 129401, 209734 to 210955, 46421 to 46874, 21635 to 24334, 132977 to 133267 and 2922116 to 2993349 in Angola, Botswana, Malawi, Namibia, Zambia, Eswatini, Lesotho, Zimbabwe and South Africa, correspondingly. The largest percentage rise in this region comes from Angola, at 18.97 percent.

4.2.4. Eastern Africa

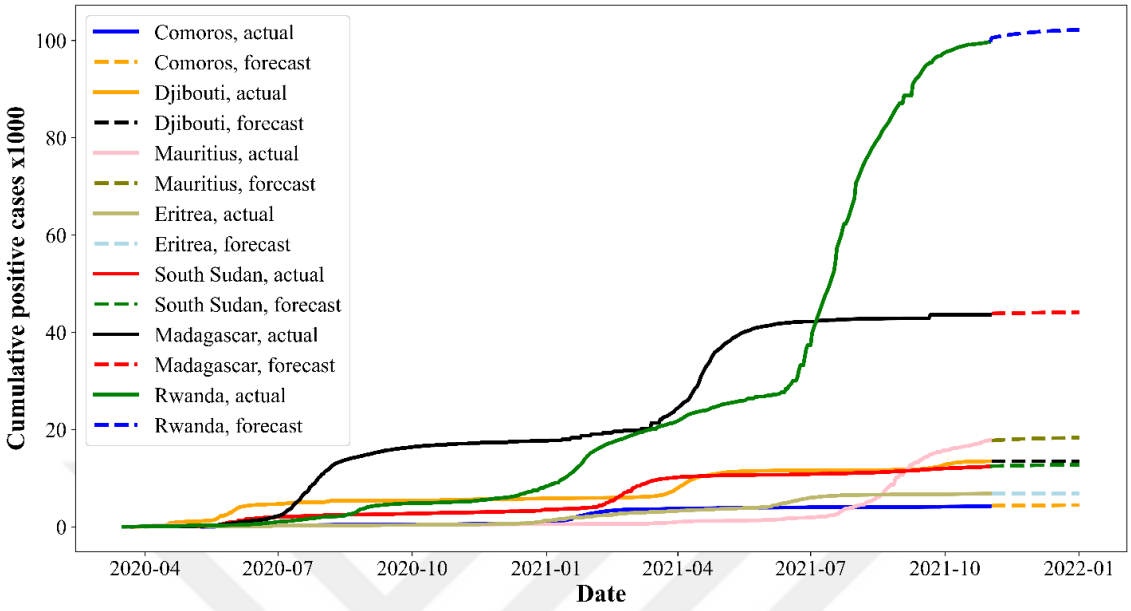

Figure 4.20. Observed and forecasted COVID-19 cases for Eastern Africa(a)

Figures 4.20 and 4.21 depict the forecasted cases in Africa's eastern area. This allowed for a thorough examination and observation of the predicted scenarios in all of the nations surveyed in this region. A graph of the actual and anticipated cases for seven nations in the Eastern African region is shown in Figure 4.19.

In most nations, the highest performing algorithm, the LSTM, has generated this projection. Depending on this estimate, there will be a modest rise in the rate of spread of COVID-19 incidents in 2 nations, Rwanda and Mauritius. Besides these 2 nations and Djibouti, which are anticipated to possess the same number of instances, the remainder of the nations are projected to have minor variations in case numbers.

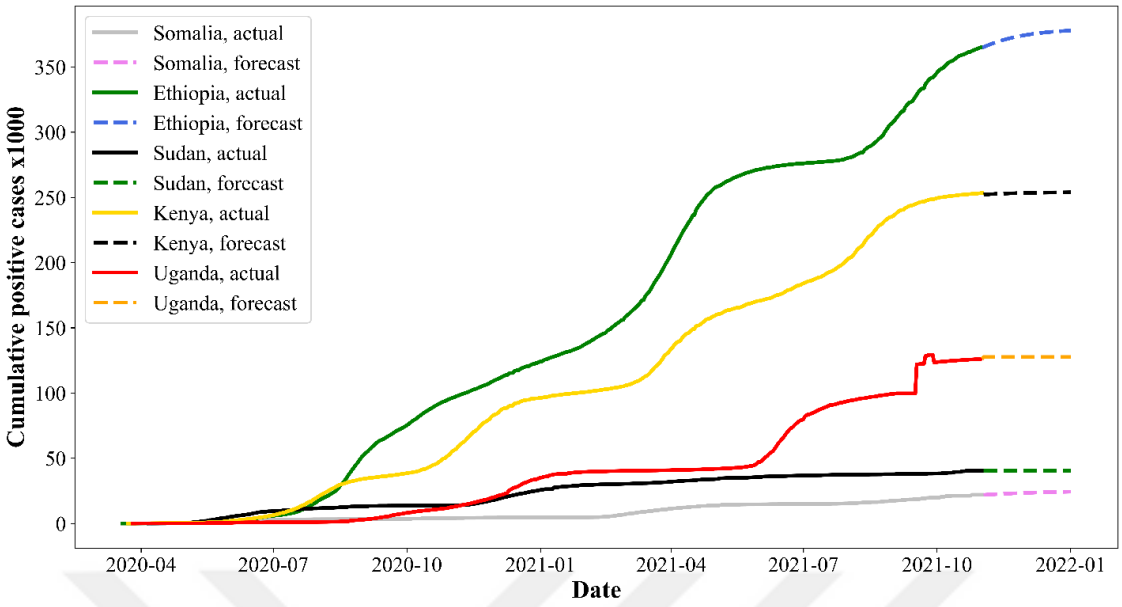

Figure 4.21. Observed and forecasted COVID-19 cases for Eastern Africa(b)

Five nations in Africa's eastern region are depicted in Figure 4.21, along with their COVID-19 cumulative positive instances. A stable number of cases is projected in Kenya, whereas an increase in Ethiopia and Somalia is expected. Both Uganda and Sudan are predicted to experience a minor increase, followed by a tiny but considerable reduction.

The instances in Djibouti are predicted to remain steady at the conclusion of the forecast timeline. The initial number of COVID-19 instances in the original dataset was 13478. That was forecasted and expected to stay the same at the conclusion of the Djibouti prediction. In Eritrea, a modest reduction from 6834 to 6820 cases is projected. By the conclusion of the forecast timeline, however, a rise is projected in the remaining nations.

Uganda, Sudan, Madagascar, Kenya, South Sudan, Somalia, Rwanda, Mauritius, Ethiopia, and the Comoros are anticipated to see an upsurge from 126236 to 127628, 40433 to 40598, 43626 to 44150, 253310 to 253901, 12410 to 12761, 21998 to 24356, 99698 to 102205, 17812 to 18297, 365167 to 377935, and 4259 to 4472. With a 10.72 percent predicted rise in the cases, Somalia is the country with the largest expected rise.

4.2.5. Western Africa

The Western African nations' predicted incidences were split into 2 groups. Six nations were plotted simultaneously in each group, as seen in figures 4.22 and 4.23. This was utilized to separate nations with similar cases for a proper examination of the forecasting phase's outcomes.

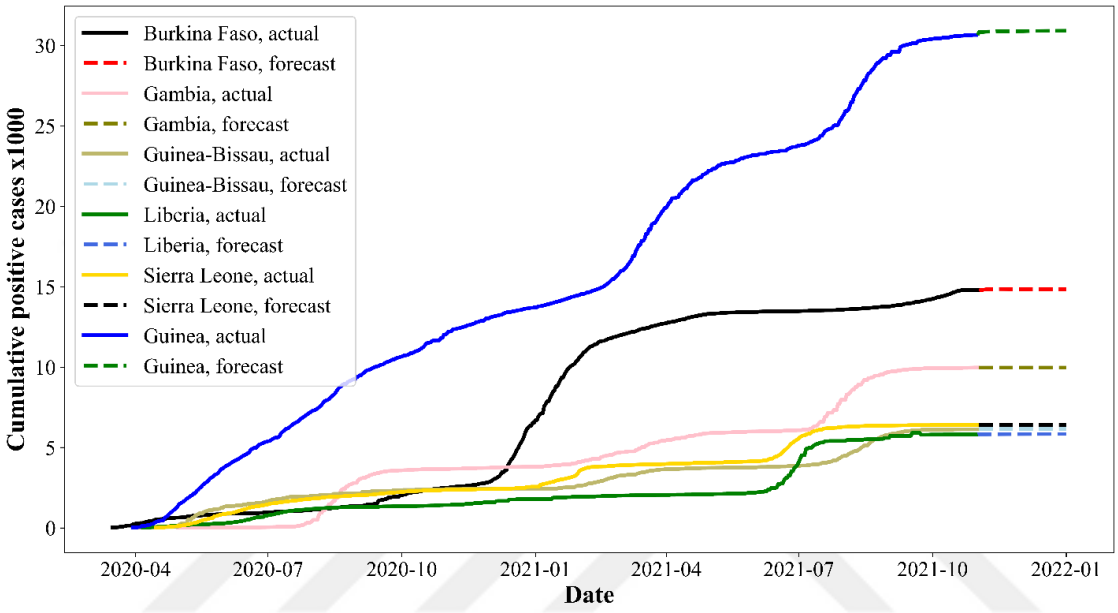

Figure 4.22. Observed and forecasted COVID-19 cases for Western Africa(a)

In figure 4.22, six nations from Africa's Western area are depicted, together with their corresponding anticipated and real cumulative cases. As per this graph, it is apparent that the projected cases in Guinea will grow somewhat, accompanied by a rather stable number of COVID-19 instances. A steady number of incidents is projected in the other 5 nations, with minor changes by the conclusion of the forecast timeline.

Because all of the nations in this graph retained their corresponding variation patterns in terms of the number of instances, it's clear that nations with more COVID-19 instance numbers before the forecasting operations, kept their higher case numbers after the forecasting. As shown in figure 4.22, nations with the highest number of actual occurrences are still likely to retain a large number of forecasted cases. Because no major reduction in the anticipated incidence is anticipated, the region remains at risk if no preventative actions are done.

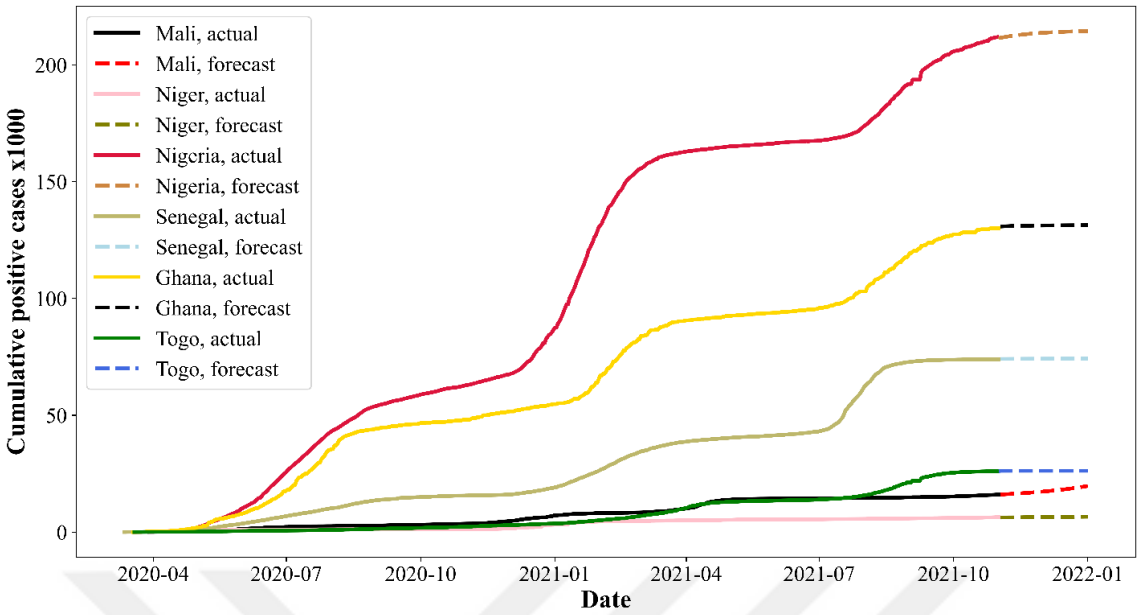

Figure 4.23. Observed and forecasted COVID-19 cases for Western Africa(b)

The other 6 nations from Africa's Western region are depicted in figure 4.23, together with anticipated and actual cases for each nation. Mali has seen a considerable rise in predicted cases, while the rest of the countries have seen a steady number of cases with minimal variations. The anticipated number of COVID-19 incidents in Gambia is projected to drop somewhat at the completion of the forecast timeline. Cases are expected to drop from 9967 to 9964 in this nation.

With the exception of the Gambia, other nations in the Western region are projected to see a rise in the number of infections. In nations like Niger, Guinea-Bissau, Guinea, Burkina Faso, Togo, Sierra Leone, Senegal, Nigeria, Mali, Liberia, and Ghana, the COVID-19 cases are projected to rise from 6366 to 6565, 6134 to 6151, 30653 to 30909, 14793 to 14848, 26079 to 26195, 6398 to 6408, 73917 to 74171, 211961 to 214433, 16074 to 19521, 5815 to 5838 and 130077 to 131344 correspondingly. By these statistics, Mali is projected to experience the highest upside percentage growth of 22.77 percent.

5. CONCLUSIONS AND SUGGESTIONS

This research used both statistical and deep learning methodologies to attain its goals, with three prediction models made up of the ARIMA, Prophet, and LSTM models. Seven performance indicators were employed in a comparison of the effectiveness of these 3 algorithms. The MSE, RMSE, MAPE, SMAPE, R2 score, NRMSE, and PSNR were among them. The best predictive model was therefore chosen to estimate future COVID-19 cases for a 61-day period. The LSTM model was the best performer in this research, whereas the Prophet algorithm was the lowest performer.

Mali is predicted to have the biggest rise in cases in the Western African region, at 22.77 percent. On the contrary, the largest predicted growth in Angola, a country in the Southern region, is 18.97 percent. Egypt, with an increase of 11.83 percent, is predicted to have the greatest rise in cases from the northern region. Somalia is anticipated to have the largest rise in cases of 10.72 percent in the eastern area. Finally, Gabon has the largest predicted rise in cases in the Central African area, at 2.81 percent.

The main challenges faced in this study comprise a lack of sufficient data given the fact that there is still not much known about the COVID-19 pandemic to train the models, especially the statistical algorithms, which demand a substantial amount of training data. There is a need for further studies including methods that employ a dynamic format of training on limited datasets to give more accurate results. There is also a need for further studies that consider a more global perspective in modeling of the models to consider the geographical aspects of the virus and other parameters such as the mutation rates of the virus and the effect of different variants.

REFERENCES

Abdulmajeed, K., Adeleke, M., & Popoola, L. (2020). ONLINE FORECASTING OF COVID-19 CASES IN NIGERIA USING LIMITED DATA. *Data in Brief*, *30*, 105683. https://doi.org/10.1016/j.dib.2020.105683.

Adeyinka, D. A., & Muhajarine, N. (2020). Time series prediction of under-five mortality rates for Nigeria: comparative analysis of artificial neural networks, Holt-Winters exponential smoothing and autoregressive integrated moving average models. *BMC Medical Research Methodology*, *20*(1). https://doi.org/10.1186/s12874-020-01159-9.

Adhikari, B., Xu, X., Ramakrishnan, N., & Prakash, B. A. (2019). EpiDeep. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. https://doi.org/10.1145/3292500.3330917.

Africa: Covid-19 Infections (National) - Humanitarian Data Exchange. (2021). Humanitarian Data Exchange. Retrieved October 1, 2021, from https://data.humdata.org/dataset/africa-covid19-infected.

African Countries by Population (2021) - Worldometer. (2021). Worldometer. Retrieved 2021, from https://www.worldometers.info/population/countries-in-africa-by-population/.

Alenezi, M. N., Al-Anzi, F. S., & Alabdulrazzaq, H. (2021). Building a sensible SIR estimation model for COVID-19 outspread in Kuwait. *Alexandria Engineering Journal*, *60*(3), 3161–3175. https://doi.org/10.1016/j.aej.2021.01.025.

Alkhayat, G., & Mehmood, R. (2021). A review and taxonomy of wind and solar energy forecasting methods based on deep learning. *Energy and AI*, *4*, 100060. https://doi.org/10.1016/j.egyai.2021.100060.

Alvarez-Ramirez, J., Soriano, A., Cisneros, M., & Suarez, R. (2003). Symmetry/anti-symmetry phase transitions in crude oil markets. *Physica A: Statistical Mechanics and Its Applications*, *322*, 583–596. https://doi.org/10.1016/s0378-4371(02)01831-9.

Arbel, N. (2021, December 7). *How LSTM networks solve the problem of vanishing gradients*. Medium. https://medium.datadriveninvestor.com/how-do-lstm-networks-solve-the-problem-of-vanishing-gradients-a6784971a577?gi=a816eea06aed.

Archontoulis, S. V., & Miguez, F. E. (2015). Nonlinear Regression Models and Applications in Agricultural Research. *Agronomy Journal*, *107*(2), 786–798. https://doi.org/10.2134/agronj2012.0506.

Attanayake, A. M. C. H., Perera, S. S. N., & Liyanage, U. P. (2020). Exponential smoothing on forecasting dengue cases in Colombo, Sri Lanka. *Journal of Science*, *11*(1), 11. https://doi.org/10.4038/jsc.v11i1.24.

Autoregressive Integrated Moving Average (ARIMA). (2021, October 12). Investopedia. Retrieved December 23, 2021, from https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp.

Blomberg, S., & Hess, G. D. (1997). Politics and exchange rate forecasts. *Journal of International Economics*, *43*(1–2), 189–205. https://doi.org/10.1016/s0022-1996(96)01466-3.

Blower, S. (2011, July 19). *The importance of including dynamic social networks when modeling epidemics of airborne infections: does increasing complexity increase accuracy?* - *BMC Medicine*. BioMed Central. https://bmcmedicine.biomedcentral.com/articles/10.1186/1741-7015-9-88.

Brauer, F., Driessche, P., Wu, J., Allen, L. J. S., van den Driessche, P., Bauch, C. T., Castillo-Chavez, C., Earn, D., Feng, Z., Lewis, M. A., & Li, J. (2008). *Mathematical Epidemiology* [E-book]. Springer Publishing.

Cai, B., & Yu, Y. (2022). Flood forecasting in urban reservoir using hybrid recurrent neural network. *Urban Climate*, *42*, 101086. https://doi.org/10.1016/j.uclim.2022.101086.

Caudron, Q., Mahmud, A. S., Metcalf, C. J. E., Gottfreðsson, M., Viboud, C., Cliff, A. D., & Grenfell, B. T. (2015). Predictability in a highly stochastic system: final size of measles epidemics in small populations. *Journal of The Royal Society Interface*, *12*(102), 20141125. https://doi.org/10.1098/rsif.2014.1125.

Cavalcante, L., Bessa, R. J., Reis, M., & Browell, J. (2016). LASSO vector autoregression structures for very short-term wind power forecasting. *Wind Energy*, *20*(4), 657–675. https://doi.org/10.1002/we.2029.

Chen, F., & Ou, T. (2011). Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry. *Expert Systems with Applications*, *38*(3), 1336–1345. https://doi.org/10.1016/j.eswa.2010.07.014.

Chen, X., Dong, Z. Y., Meng, K., Xu, Y., Wong, K. P., & Ngan, H. W. (2012). Electricity Price Forecasting With Extreme Learning Machine and Bootstrapping. *IEEE Transactions on Power Systems*, *27*(4), 2055–2062. https://doi.org/10.1109/tpwrs.2012.2190627.

Chintalapudi, N., Angeloni, U., Battineni, G., di Canio, M., Marotta, C., Rezza, G., Sagaro, G. G., Silenzi, A., & Amenta, F. (2022). LASSO Regression Modeling on Prediction of Medical Terms among Seafarers' Health Documents Using Tidy Text Mining. *Bioengineering*, *9*(3), 124. https://doi.org/10.3390/bioengineering9030124.

Choubin, B., Khalighi-Sigaroodi, S., Malekian, A., & Kişi, Z. (2016). Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. *Hydrological Sciences Journal*, *61*(6), 1001–1009. https://doi.org/10.1080/02626667.2014.966721.

Chowell, G., Tariq, A., & Hyman, J. M. (2019, August 22). *A novel sub-epidemic modeling framework for short-term forecasting epidemic waves - BMC Medicine*. SpringerLink. https://link.springer.com/article/10.1186/s12916-019-1406-6?error=cookies_not_supported&code=402e970a-0940-463c-bb74-fcf6a299066f.

Corporate Finance Institute. (2022, January 19). *Autoregressive Integrated Moving Average (ARIMA)*. Retrieved January 19, 2022, from https://corporatefinanceinstitute.com/resources/knowledge/other/autoregressive-integrated-moving-average-arima/.

Dai, Z., & Zhu, H. (2020). Stock return predictability from a mixed model perspective. *Pacific-Basin Finance Journal*, *60*, 101267. https://doi.org/10.1016/j.pacfin.2020.101267.

Das, J., Manikanta, V., Nikhil Teja, K., & Umamahesh, N. V. (2022). Two decades of ensemble flood forecasting: a state-of-the-art on past developments, present applications and future opportunities. *Hydrological Sciences Journal*, *67*(3), 477–493. https://doi.org/10.1080/02626667.2021.2023157.

de Myttenaere, A., Golden, B., le Grand, B., & Rossi, F. (2016). Mean Absolute Percentage Error for regression models. *Neurocomputing*, *192*, 38–48. https://doi.org/10.1016/j.neucom.2015.12.114.

de Oliveira, B. N., Valk, M., & Marcondes Filho, D. (2022). Fault detection and diagnosis of batch process dynamics using ARMA-based control charts. *Journal of Process Control*, *111*, 46–58. https://doi.org/10.1016/j.jprocont.2022.01.005.

Ediger, V., & Akar, S. (2007). ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy*, *35*(3), 1701–1708. https://doi.org/10.1016/j.enpol.2006.05.009.

Efimov, D., & Ushirobira, R. (2021). On an interval prediction of COVID-19 development based on a SEIR epidemic model. *Annual Reviews in Control*, *51*, 477–487. https://doi.org/10.1016/j.arcontrol.2021.01.006.

Fang, D., Zhang, X., Yu, Q., Jin, T. C., & Tian, L. (2018). A novel method for carbon dioxide emission forecasting based on improved Gaussian processes regression. *Journal of Cleaner Production*, *173*, 143–150. https://doi.org/10.1016/j.jclepro.2017.05.102.

Fang, T., & Lahdelma, R. (2016). Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. *Applied Energy*, *179*, 544–552. https://doi.org/10.1016/j.apenergy.2016.06.133.

Faust, J., Rogers, J. H., & H. Wright, J. (2003). Exchange rate forecasting: the errors we've really made. *Journal of International Economics*, *60*(1), 35–59. https://doi.org/10.1016/s0022-1996(02)00058-2.

Feng, X., Ma, G., Su, S. F., Huang, C., Boswell, M. K., & Xue, P. (2020). A multi-layer perceptron approach for accelerated wave forecasting in Lake Michigan. *Ocean Engineering*, *211*, 107526. https://doi.org/10.1016/j.oceaneng.2020.107526.

Frost, J. (2021, May 18). *Exponential Smoothing for Time Series Forecasting*. Statistics By Jim. Retrieved April 25, 2022, from https://statisticsbyjim.com/time-series/exponential-smoothing-time-series-forecasting/.

Fumi, A., Pepe, A., Scarabotti, L., & Schiraldi, M. M. (2013). Fourier Analysis for Demand Forecasting in a Fashion Company. *International Journal of Engineering Business Management*, *5*, 30. https://doi.org/10.5772/56839.

Funk, S., Camacho, A., Kucharski, A. J., Eggo, R. M., & Edmunds, W. J. (2018). Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. *Epidemics*, *22*, 56–61. https://doi.org/10.1016/j.epidem.2016.11.003.

Gebretensae, Y. A., & Asmelash, D. (2021). Trend Analysis and Forecasting the Spread of COVID-19 Pandemic in Ethiopia Using Box–Jenkins Modeling Procedure. *International Journal of General Medicine*, *Volume 14*, 1485–1498. https://doi.org/10.2147/ijgm.s306250.

Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., di Filippo, A., di Matteo, A., & Colaneri, M. (2020, March 22). *A SIDARTHE Model of COVID-19 Epidemic in Italy*. arXiv.Org. https://arxiv.org/abs/2003.09861.

Glen, S. (2021, January 12). *Weighted Least Squares: Simple Definition, Advantages & Disadvantages*. Statistics How To. Retrieved April 23, 2022, from https://www.statisticshowto.com/weighted-least-squares/.

Gomes, P., & Castro, R. (2012). Wind Speed and Wind Power Forecasting using Statistical Models: AutoRegressive Moving Average (ARMA) and Artificial Neural Networks (ANN). *International Journal of Sustainable Energy Development*, *1*(2), 41–50. https://doi.org/10.20533/ijsed.2046.3707.2012.0007.

Guan, P., Wu, W., & Huang, D. (2018). Trends of reported human brucellosis cases in mainland China from 2007 to 2017: an exponential smoothing time series analysis. *Environmental Health and Preventive Medicine*, *23*(1). https://doi.org/10.1186/s12199-018-0712-5.

Hamel, P., & Eck, D. (2010). Learning features from music audio with deep belief networks. *ISMIR*, *10*, 339–344.

Han, Z., Zhao, J., Leung, H., Ma, K. F., & Wang, W. (2021). A Review of Deep Learning Models for Time Series Prediction. *IEEE Sensors Journal*, *21*(6), 7833–7848. https://doi.org/10.1109/jsen.2019.2923982.

Hanck, C., Arnold, M., Gerber, A., & Schmelzer, M. (2021, October 6). *Introduction to Econometrics with R*. Econometrics-with-r. Retrieved April 21, 2022, from https://www.econometrics-with-r.org/.

Herwartz, H. (2017). Stock return prediction under GARCH — An empirical assessment. *International Journal of Forecasting*, *33*(3), 569–580. https://doi.org/10.1016/j.ijforecast.2017.01.002.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*(1), 5–10. https://doi.org/10.1016/j.ijforecast.2003.09.015.

Hssayeni, M. D., Chala, A., Dev, R., Xu, L., Shaw, J., Furht, B., & Ghoraani, B. (2021). The forecast of COVID-19 spread risk at the county level. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00491-1.

Huang, C. J., Chen, Y. H., Ma, Y., & Kuo, P. H. (2020). Multiple-Input Deep Convolutional Neural Network Model for COVID-19 Forecasting in China. *medRxiv*. https://doi.org/10.1101/2020.03.23.20041608.

Hutchinson, B., Deng, L., & Yu, D. (2013). Tensor Deep Stacking Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1944–1957. https://doi.org/10.1109/tpami.2012.268.

Hyndman, R. J., & Athanasopoulos, G. (2022). *Forecasting: principles and practice* (2nd ed.) [E-book]. OTexts.

Hyndman, R., Koehler, A. B., Ord, K. J., & Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach (Springer Series in Statistics)* (2008th ed.). Springer.

Kamp, C. (2012, August 31). *Predicting epidemics on weighted networks*. arXiv.Org. https://arxiv.org/abs/1208.6497.

Kantar, Y. M. (2015). GENERALIZED LEAST SQUARES AND WEIGHTED LEAST SQUARES ESTIMATION METHODS FOR DISTRIBUTIONAL PARAMETERS. *REVSTAT – Statistical Journa*, *13*(3), 263–282. https://www.ine.pt/revstat/autores/pdf/rs150306.pdf.

Kırbaş, S., Sözen, A., Tuncer, A. D., & Kazancıoğlu, F. I. (2020). Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches. *Chaos, Solitons & Fractals*, *138*, 110015. https://doi.org/10.1016/j.chaos.2020.110015.

Kourav, A. K., Sharma, S., & Tiwari, V. (2018). Wavelet Transform Algorithms. *Advances in Systems Analysis, Software Engineering, and High Performance Computing*, 177–192. https://doi.org/10.4018/978-1-5225-3870-7.ch011.

Li, J., Tsiakas, I., & Wang, W. (2014). Predicting Exchange Rates Out of Sample: Can Economic Fundamentals Beat the Random Walk? *Journal of Financial Econometrics*, *13*(2), 293–341. https://doi.org/10.1093/jjfinec/nbu007.

Li, T., Hua, M., & Wu, X. (2020). A Hybrid CNN-LSTM Model for Forecasting Particulate Matter (PM2.5). *IEEE Access*, *8*, 26933–26940. https://doi.org/10.1109/access.2020.2971348.

Li, X., Huang, X., Deng, X., & Zhu, S. (2014). Enhancing quantitative intra-day stock return prediction by integrating both market news and stock prices information. *Neurocomputing*, *142*, 228–238. https://doi.org/10.1016/j.neucom.2014.04.043.

Liang, C. W., Ku, C. K., & Liang, J. J. (2013). A new scale-free network model for simulating and predicting epidemics. *Journal of Theoretical Biology*, *317*, 11–19. https://doi.org/10.1016/j.jtbi.2012.09.020.

Lin, T., & Tsai, C. C. L. (2015). A Simple Linear Regression Approach to Modeling and Forecasting Mortality Rates. *Journal of Forecasting*, *34*(7), 543–559. https://doi.org/10.1002/for.2353.

Liu, H., Mi, X. W., & Li, Y. F. (2018). Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. *Energy Conversion and Management*, *156*, 498–514. https://doi.org/10.1016/j.enconman.2017.11.053.

Livieris, I. E., Pintelas, E., & Pintelas, P. (2020). A CNN–LSTM model for gold price time-series forecasting. *Neural Computing and Applications*, *32*(23), 17351–17360. https://doi.org/10.1007/s00521-020-04867-x.

Mahajan, A., Sivadas, N. A., & Solanki, R. (2020). An epidemic model SIPHERD and its application for prediction of the spread of COVID-19 infection in India. *Chaos, Solitons & Fractals*, *140*, 110156. https://doi.org/10.1016/j.chaos.2020.110156.

Majid, R., & Mir, S. A. (2018). Advances in Statistical Forecasting Methods: An Overview. *Economic Affairs*, *63*(4). https://doi.org/10.30954/0424-2513.4.2018.5.

Mapa, D. S. (2004). A Forecast Comparison of Financial Volatility Models: GARCH (1,1) is not Enough. *Munich Personal RePEc Archive*, *531*, 1–10. https://mpra.ub.uni-muenchen.de/21028/.

Marzouk, M., Elshaboury, N., Abdel-Latif, A., & Azab, S. (2021). Deep learning model for forecasting COVID-19 outbreak in Egypt. *Process Safety and Environmental Protection*, *153*, 363–375. https://doi.org/10.1016/j.psep.2021.07.034.

Mathworks. (2021). *Autocorrelation and Partial Autocorrelation - MATLAB & Simulink*. Mathworks. Retrieved December 6, 2021, from https://www.mathworks.com/help/econ/autocorrelation-and-partial-autocorrelation.html.

McKenzie, E., & Gardner, E. S. (2010). Damped trend exponential smoothing: A modelling viewpoint. *International Journal of Forecasting*, *26*(4), 661–665. https://doi.org/10.1016/j.ijforecast.2009.07.001.

MEAN ABSOLUTE PERCENTAGE ERROR (MAPE). (2000). *Encyclopedia of Production and Manufacturing Management*, 462. https://doi.org/10.1007/1-4020-0612-8_580.

Meenal, R., Binu, D., Ramya, K. C., Michael, P. A., Vinoth Kumar, K., Rajasekaran, E., & Sangeetha, B. (2022). Weather Forecasting for Renewable Energy System: A Review. *Archives of Computational Methods in Engineering*. https://doi.org/10.1007/s11831-021-09695-3.

Meyers, L. A., Newman, M., & Pourbohloul, B. (2006). Predicting epidemics on directed contact networks. *Journal of Theoretical Biology*, *240*(3), 400–418. https://doi.org/10.1016/j.jtbi.2005.10.004.

Mohamed, A. R., Dahl, G., & Hinton, G. (2009). Deep belief networks for phone recognition. *Nips Workshop on Deep Learning for Speech Recognition and Related Applications*, *1*(9), 39.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting (Wiley Series in Probability and Statistics)* (2nd ed.). Wiley-Interscience.

Moore, P. J., & Little, M. A. (2014). Enhancements to a method of analogues forecasting algorithm. *The 2014 International Symposium on Nonlinear Theory and Its Applications*, 317–320. https://doi.org/10.34385/proc.46.B2L-B3.

Murugan Bhagavathi, S., Thavasimuthu, A., Murugesan, A., George Rajendran, C. P. L., A, V., Raja, L., & Thavasimuthu, R. (2021). Weather forecasting and prediction using hybrid C5.0 machine learning algorithm. *International Journal of Communication Systems*, *34*(10). https://doi.org/10.1002/dac.4805.

Nenni, M. E., Giustiniano, L., & Pirolo, L. (2013). Demand Forecasting in the Fashion Industry: A Review. *International Journal of Engineering Business Management*, *5*, 37. https://doi.org/10.5772/56840.

Noureen, S., Atique, S., Roy, V., & Bayne, S. (2019). Analysis and application of seasonal ARIMA model in Energy Demand Forecasting: A case study of small scale agricultural load. *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*. https://doi.org/10.1109/mwscas.2019.8885349.

Nsoesie, E. O., Brownstein, J. S., Ramakrishnan, N., & Marathe, M. V. (2013). A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza and Other Respiratory Viruses*, *8*(3), 309–316. https://doi.org/10.1111/irv.12226.

Oladunni, T., Denis, M., Ososanya, E., & Barry, A. (2021). Exponential Smoothing Forecast of African Americans' COVID-19 Fatalities. *2021 2nd International Conference on Computing and Data Science (CDS)*. https://doi.org/10.1109/cds52072.2021.00086.

Pal, R., Sekh, A. A., Kar, S., & Prasad, D. K. (2020). Neural Network Based Country Wise Risk Prediction of COVID-19. *Applied Sciences*, *10*(18), 6448. https://doi.org/10.3390/app10186448.

Panda, C., & Narasimhan, V. (2007). Forecasting exchange rate better with artificial neural network. *Journal of Policy Modeling*, *29*(2), 227–236. https://doi.org/10.1016/j.jpolmod.2006.01.005.

Pandey, G., Chaudhary, P., Gupta, R., & Pal, S. (2020, April 1). *SEIR and Regression Model based COVID-19 outbreak predictions in India*. arXiv.Org. https://arxiv.org/abs/2004.00958.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., de Baets, S., Dokumentov, A., . . . Ziel, F. (2022). Forecasting: theory and practice. *International Journal of Forecasting*. https://doi.org/10.1016/j.ijforecast.2021.11.001.

Piadeh, F., Behzadian, K., & Alani, A. M. (2022). A critical review of real-time modelling of flood forecasting in urban drainage systems. *Journal of Hydrology*, *607*, 127476. https://doi.org/10.1016/j.jhydrol.2022.127476.

Pinson, P., Chevallier, C., & Kariniotakis, G. N. (2007, August 1). *Trading Wind Generation From Short-Term Probabilistic Forecasts of Wind Power*. IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/4282048?casa_token=Q98jiJ7C9AQAAAAA:7BVvzegQqfwvlUMbo_xfPhM2iKQpOcQYYK17A7pwywenEm8m2H5bTsBETzY53kswxcX1p4UkhI.

Ribeiro, M. H. D. M., da Silva, R. G., Mariani, V. C., & Coelho, L. D. S. (2020a). Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos, Solitons & Fractals*, *135*, 109853. https://doi.org/10.1016/j.chaos.2020.109853.

Ribeiro, M. H. D. M., da Silva, R. G., Mariani, V. C., & Coelho, L. D. S. (2020b). Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos, Solitons & Fractals*, *135*, 109853. https://doi.org/10.1016/j.chaos.2020.109853.

Salehinejad, H., Sankar, S., Barfett, J., Colak, E., & Valaee, S. (2017, December 29). *Recent Advances in Recurrent Neural Networks*. arXiv.Org. https://arxiv.org/abs/1801.01078.

Seneviratna, D., & Rathnayaka, R. K. T. (2022). Hybrid grey exponential smoothing approach for predicting transmission dynamics of the COVID-19 outbreak in Sri Lanka. *Grey Systems: Theory and Application*. https://doi.org/10.1108/gs-06-2021-0085.

Shahid, F., Zameer, A., & Muneeb, M. (2020). Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons & Fractals*, *140*, 110212. https://doi.org/10.1016/j.chaos.2020.110212.

Shastri, S., Singh, K., Kumar, S., Kour, P., & Mansotra, V. (2020). Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study. *Chaos, Solitons & Fractals*, *140*, 110227. https://doi.org/10.1016/j.chaos.2020.110227.

Shen, G., Tan, Q., Zhang, H., Zeng, P., & Xu, J. (2018). Deep Learning with Gated Recurrent Unit Networks for Financial Sequence Predictions. *Procedia Computer Science*, *131*, 895–903. https://doi.org/10.1016/j.procs.2018.04.298.

Shenstone, L., & Hyndman, R. J. (2005). Stochastic models underlying Croston's method for intermittent demand forecasting. *Journal of Forecasting*, *24*(6), 389–402. https://doi.org/10.1002/for.963.

Singh, R. K., Rani, M., Bhagavathula, A. S., Sah, R., Rodriguez-Morales, A. J., Kalita, H., Nanda, C., Sharma, S., Sharma, Y. D., Rabaan, A. A., Rahmani, J., & Kumar, P. (2020). Prediction of the COVID-19 Pandemic for the Top 15 Affected Countries: Advanced Autoregressive Integrated Moving Average (ARIMA) Model. *JMIR Public Health and Surveillance*, *6*(2), e19115. https://doi.org/10.2196/19115.

Soyiri, I. N., & Reidpath, D. D. (2012). An overview of health forecasting. *Environmental Health and Preventive Medicine*, *18*(1), 1–9. https://doi.org/10.1007/s12199-012-0294-6.

Srinivas, K., Kavitha Rani, B., Vara Prasad Rao, M., Kumar Patra, R., Madhukar, G., & Mahendar, A. (2020). Prediction Of Heart Disease Using Hybrid Linear Regression. *European Journal of Molecular & Clinical Medicine*, *07*(05), 2515–8260. https://ejmcm.com/article_3785_56dea6128008c7563d95cb35313a0908.pdf.

Taylor, A. M., & Taylor, M. P. (2004). The Purchasing Power Parity Debate. *Journal of Economic Perspectives*, *18*(4), 135–158. https://doi.org/10.1257/0895330042632744.

Taylor, S. J., & Letham, B. (2017). Forecasting at scale. *PeerJ Preprints*. https://doi.org/10.7287/peerj.preprints.3190v2.

Thongpeth, W., Lim, A., Wongpairin, A., Thongpeth, T., & Chaimontree, S. (2021). Comparison of linear, penalized linear and machine learning models predicting hospital visit costs from chronic disease in Thailand. *Informatics in Medicine Unlocked*, *26*, 100769. https://doi.org/10.1016/j.imu.2021.100769.

Tiwari, S., Sabzehgar, R., & Rasouli, M. (2018, June 1). *Short Term Solar Irradiance Forecast Using Numerical Weather Prediction (NWP) with Gradient Boost Regression*. IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8447751?casa_token=7ON2sVHz86sAAAAA:v8XBbj8kmbTT9GnAzwRontp7NWw1UUjZjbf3CelyoZE5DUPP02dCaZ-nkrX5lI1EEdgbgNLV3VY.

Viboud, C. (2003). Prediction of the Spread of Influenza Epidemics by the Method of Analogues. *American Journal of Epidemiology*, *158*(10), 996–1006. https://doi.org/10.1093/aje/kwg239.

Victor Devadoss, A., & Antony Alphonnse Ligori, T. (2013). Forecasting of Stock Prices Using Multi Layer Perceptron. *International Journal of Web Technology*, *002*(002), 52–58. https://doi.org/10.20894/ijwt.104.002.002.006.

Wan, C., Xu, Z., Pinson, P., Dong, Z. Y., & Wong, K. P. (2014). Probabilistic Forecasting of Wind Power Generation Using Extreme Learning Machine. *IEEE Transactions on Power Systems*, *29*(3), 1033–1044. https://doi.org/10.1109/tpwrs.2013.2287871.

Wang, H., Chien, C., & Liu, C. (2005). Demand Forecasting Using Bayesian Experiment with Non-homogenous Poisson Process Model. *International Journal of Operations Research*, *2*(1), 21–29. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.401.9597&rep=rep1&type=pdf.

Wang, H., Wang, G., Li, G., Peng, J., & Liu, Y. (2016). Deep belief network based deterministic and probabilistic wind speed forecasting approach. *Applied Energy*, *182*, 80–93. https://doi.org/10.1016/j.apenergy.2016.08.108.

Wang, H. Z., Li, G. Q., Wang, G. B., Peng, J. C., Jiang, H., & Liu, Y. T. (2017). Deep learning based ensemble approach for probabilistic wind power forecasting. *Applied Energy*, *188*, 56–70. https://doi.org/10.1016/j.apenergy.2016.11.111.

Wang, J. (2017). *Development of Satellite-Assisted Forecasting System for Oyster Norovirus Outbreaks*. LSU Digital Commons. https://digitalcommons.lsu.edu/gradschool_dissertations/1570/.

Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X., & Wang, F. Y. (2017). Generative adversarial networks: introduction and outlook. *IEEE/CAA Journal of Automatica Sinica*, *4*(4), 588–598. https://doi.org/10.1109/jas.2017.7510583.

Wang, K., Li, K., Zhou, L., Hu, Y., Cheng, Z., Liu, J., & Chen, C. (2019). Multiple convolutional neural networks for multivariate time series prediction. *Neurocomputing*, *360*, 107–119. https://doi.org/10.1016/j.neucom.2019.05.023.

Wang, L., Zhang, Z., & Chen, J. (2017). Short-Term Electricity Price Forecasting With Stacked Denoising Autoencoders. *IEEE Transactions on Power Systems*, *32*(4), 2673–2681. https://doi.org/10.1109/tpwrs.2016.2628873.

Wang, P., Zheng, X., Li, J., & Zhu, B. (2020). Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. *Chaos, Solitons & Fractals*, *139*, 110058. https://doi.org/10.1016/j.chaos.2020.110058.

Wang, Y., Hug, G., Liu, Z., & Zhang, N. (2020). Modeling load forecast uncertainty using generative adversarial networks. *Electric Power Systems Research*, *189*, 106732. https://doi.org/10.1016/j.epsr.2020.106732.

Wang, Y., Wang, J., Zhao, G., & Dong, Y. (2012). Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China. *Energy Policy*, *48*, 284–294. https://doi.org/10.1016/j.enpol.2012.05.026.

Weisberg, S. (2013). *Applied Linear Regression* (4th ed.) [E-book]. Wiley.

Wikipedia contributors. (2022a, April 17). *Weighted least squares*. Wikipedia. Retrieved April 23, 2022, from

https://en.wikipedia.org/wiki/Weighted_least_squares#:%7E:text=Weighted%20least%20squares%20(WLS)%2C,specialization%20of%20generalized%20least%20squares.

Wikipedia contributors. (2022b, May 7). *Regions of the African Union*. Wikipedia. Retrieved 2021, from https://en.wikipedia.org/wiki/Regions_of_the_African_Union.

Willemain, T. R., Smart, C. N., & Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting*, *20*(3), 375–387. https://doi.org/10.1016/s0169-2070(03)00013-x.

Wright, J. H. (2008). Bayesian Model Averaging and exchange rate forecasts. *Journal of Econometrics*, *146*(2), 329–341. https://doi.org/10.1016/j.jeconom.2008.08.012.

Wu, W., Guo, J., An, S., Guan, P., Ren, Y., Xia, L., & Zhou, B. (2015). Comparison of Two Hybrid Models for Forecasting the Incidence of Hemorrhagic Fever with Renal Syndrome in Jiangsu Province, China. *PLOS ONE*, *10*(8), e0135492. https://doi.org/10.1371/journal.pone.0135492.

Xie, W., Yu, L., Xu, S., & Wang, S. (2006). *A New Method for Crude Oil Price Forecasting Based on Support Vector Machines*. SpringerLink. https://link.springer.com/chapter/10.1007/11758549_63?error=cookies_not_supported&code=3aa6ea66-b591-45f2-8aa8-fabd387a3b5d.

Xishuang Dong, Lijun Qian, & Lei Huang. (2017). Short-term load forecasting in smart grid: A combined CNN and K-means clustering approach. *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*. https://doi.org/10.1109/bigcomp.2017.7881726.

Yang, W., Karspeck, A., & Shaman, J. (2014). Comparison of Filtering Methods for the Modeling and Retrospective Forecasting of Influenza Epidemics. *PLoS Computational Biology*, *10*(4), e1003583. https://doi.org/10.1371/journal.pcbi.1003583.

Yang, Z., Zeng, Z., Wang, K., Wong, S. S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li, F., Luo, S., . . . He, J. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, *12*(3), 165–174. https://doi.org/10.21037/jtd.2020.02.64.

Yasar, A., Bilgili, M., & Simsek, E. (2012). Water Demand Forecasting Based on Stepwise Multiple Nonlinear Regression Analysis. *Arabian Journal for Science and Engineering*, *37*(8), 2333–2341. https://doi.org/10.1007/s13369-012-0309-z.

Yu, C. S., Chang, S. S., Chang, T. H., Wu, J. L., Lin, Y. J., Chien, H. F., & Chen, R. J. (2021). A COVID-19 Pandemic Artificial Intelligence–Based System With Deep Learning Forecasting and Automatic Statistical Data Acquisition: Development and Implementation Study. *Journal of Medical Internet Research*, *23*(5), e27806. https://doi.org/10.2196/27806.

Zeroual, A., Harrou, F., Dairi, A., & Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study. *Chaos, Solitons & Fractals*, *140*, 110121. https://doi.org/10.1016/j.chaos.2020.110121.

Zhang, C. Y., Chen, C. L. P., Gan, M., & Chen, L. (2015). Predictive Deep Boltzmann Machine for Multiperiod Wind Speed Forecasting. *IEEE Transactions on Sustainable Energy*, *6*(4), 1416–1425. https://doi.org/10.1109/tste.2015.2434387.

Zhang, P., & Ci, B. (2020). Deep belief network for gold price forecasting. *Resources Policy*, *69*, 101806. https://doi.org/10.1016/j.resourpol.2020.101806.

Zhao, X. (2017, November 16). *Forecasting carbon dioxide emissions based on a hybrid of mixed data sampling regression model and back propagation neural network in the USA*. SpringerLink. https://link.springer.com/article/10.1007/s11356-017-0642-6?error=cookies_not_supported&code=7c526854-a184-4525-8b2a-81c562770db6#Abs1.

Zhao, X., & Du, D. (2015). Forecasting carbon dioxide emissions. *Journal of Environmental Management*, *160*, 39–44. https://doi.org/10.1016/j.jenvman.2015.06.002.

Zoabi, Y., Deri-Rozov, S., & Shomron, N. (2021). Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *Npj Digital Medicine*, *4*(1). https://doi.org/10.1038/s41746-020-00372-6.

GAZİLİ OLMAK AYRICALIKTIR…