Building Univariate Time Series Forecasting Models for Network Traffic by Evaluating Statistical and Machine Leaning Techniques: The Case of Ethio Telecom Broadband VSAT Hub Networks

BY: ERMIAS NIGUSSIE

SUPERVISOR: EPHREM TESHALE (PhD)

A Thesis Submitted to
School of Electrical and Computer Engineering
Addis Ababa Institute of Technology

in Partial Fulfillment of the Requirements for the Degree of Master of Science
(Telecommunication Engineering)

Addis Ababa University
Addis Ababa, Ethiopia

January 18, 2022

Declaration

I, the undersigned, declare that the thesis comprises my own work in compliance with internationally accepted practices; I have fully acknowledged and referred all materials used in this thesis work.

Ermias Nigussie

____________________ ____________________

Name Signature

Addis Ababa University

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

This is to certify that the thesis prepared by **Ermias Nigussie**, entitled *Building Univariate Time Series Forecasting Models for Network Traffic by Evaluating Statistical and Machine Leaning Techniques: The Case of Ethio Telecom Broadband VSAT Hub Networks* and submitted in partial fulfillment of the requirements for the degree of Master of Science (Telecommunication Engineering) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Examiner 1	Surafel Lemma (PhD)	Signature	________	Date	______
Examiner 2	Rosa Tsegaye (PhD)	Signature	________	Date	______
Supervisor	Ephrem Teshale (PhD)	Signature	________	Date	______

____

Dean, School of Electrical and Computer Engineering

Dedicated to my family.

ABSTRACT

Network traffic congestion is the major challenge in telecom service providers where usually they use a limited resource to deliver the services to their customers. That leads to network performance and Quality of Services (QoS) degradation so that do not meet customer satisfaction. The Very Small Aperture Terminal (VSAT) network in Ethio Telecom delivers broadband services through satellite with a limited capacity. Therefore, this thesis aims to study the VSAT network traffic patterns to propose the traffic forecasting model. That will be used as a solution to enhance the network resources based on the prediction of the future traffic demand and as input for network planning and optimization works.

In this study, the VSAT data traffic recorded for one year from 01-Mar-2020 to 28-Feb-2021 was collected. The dataset is used for data preprocessing, statistical analysis, model training and testing. All the tasks are performed by Python software. In addition, existing time series forecasting methods are selected from statistical and machine learning models that includes the Exponential Smoothing Methods (ESM), Autoregressive Integrated Moving Averages (ARIMA), Seasonal ARIMA (SARIMA), Artificial Neural Network (ANN) variants Multilayer Perceptron (MLP), Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM). The forecasting accuracy metric of Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are used to evaluate the forecasting performance of the models. These are applied to examine and choose the model having the minimum forecasting errors. As a result, the RNN model is identified the best model and improved the forecasting performance by 44.94% than the Triple Exponential Smoothing (TES) model which is a variant of ESM. Therefore, the RNN model is proposed to Ethio Telecom for future network planning and optimization to VSAT networks.

KEYWORDS

VSAT, Network data traffic, Forecasting, ESM, ARIMA, SARIMA, ANN.

ACKNOWLEDGMENTS

First and foremost, I thank and praises to the almighty God, for his loving kindness and great mercy besides helping me to complete this study!

Next, I would like to express my special thanks and appreciation to my adviser Dr. Ephrem Teshale, for his excellent comments, guidance, advice and encouragement. I also thank my evaluators, Dr. Surafel and Dr. Rosa for their suggestions and feedbacks during thesis progress seminars that helped me to go through my thesis work.

Moreover, I would like to thank Ethio Telecom, my employer, for giving me the opportunity to attend my postgraduate program in Addis Ababa University. I also would like sincerely to thank Ethio Telecom staffs Mr. Michael Melaku for providing me VSAT traffic data used as an input for this thesis; Mr. Demissie Zewide and Mr. Abdulhafiz Luelseged who helped me in providing necessary information and materials, including telecommunication engineering students for their direct and indirect support.

Last but not least, I want to express my deepest gratitude to my beloved spouse Mrs. Bezawit Belayneh and to my little baby girl Mariamawit, including my relatives and all friends for their love, support and patience.

This page is intentionally left blank.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ACF	Autocorrelation Function
ADF	Augmented Dickey Fuller
ANN	Artificial Neural Networks
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Averages
ARMA	Autoregressive Moving Average
ASR	Aggregated Service Router
CSV	Comma Separated Values
DES	Double Exponential Smoothing
DVB-S2	Digital Video Broadcasting - Satellite Second Generation
ESM	Exponential Smoothing Methods
IQR	Interquartile Range
IP	Internet Protocol
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
LR	Linear Regression
LSTM	Long Short Term Memory
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MLP	Multilayer Perceptron
MSE	Mean Square Error
NN	Neural Network
PACF	Partial Autocorrelation Function
QoS	Quality of Service
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SARIMA	Seasonal ARIMA
SES	Single Exponential Smoothing

TES Triple Exponential Smoothing

UMTS Universal Mobile Telecommunications System

VSAT Very Small Aperture Terminal

1 INTRODUCTION

Nowadays, telecom providers have to deal with continuously growing traffic demands. The demand of the users, due to the evolving of applications and services, the traffic growth becomes very high. It is also crucial to guarantee the demand of bandwidth with the required Quality of Service (QoS). Most telecom operators deploy broadband networks to deliver various services such as data, video, and voice to their customers [1].

Ethio Telecom is a state-owned telecom operator and runs different networks to deliver fixed and wireless telecom services to a large number of subscribers. Currently, the demand for broadband service is remarkably increasing and more than 375,000 subscribers utilize the service. The operator offers broadband services through fixed-wired with access speed starting from 1Mbps. In addition, it delivers up to 5Mbps through fixed-wireless Very Small Aperture Terminal (VSAT) network. Ethio Telecom owns the VSAT network to provide service via satellite as an alternative solution where there is no terrestrial network or as a backup. The VSAT hub offers services like a backhauling Universal Mobile Telecommunications System (UMTS) cellular network, rural telecom connectivity, schools distance learning, hosting corporate and government administrations. Currently, the networks accommodate several users with a limited satellite bandwidth capacity. There are high spikes or bursty and unbalanced traffic management on Internet Protocol (IP) gateways. Besides, as shown in Figure 1.1, high data users are now emerging and the networks are experiencing high network congestion [2] [3].

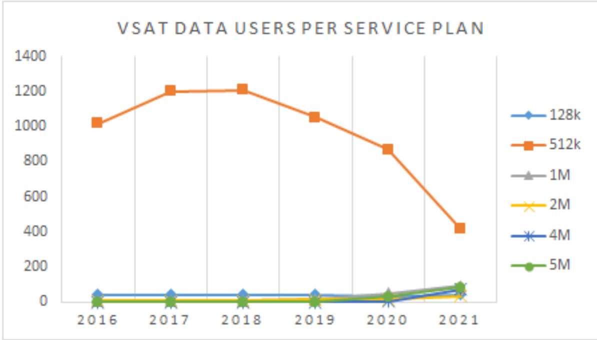

Figure 1.1: Broadband VSAT users [3].

The traffic generated in broadband services has various statistical characteristics and QoS that the network requires. The broadband traffic pattern is complex so it requires modeling and analysis [4]. Moreover, the analysis needs the collection of greater traffic volume as its pattern changes over time [5] [6].

In network traffic modeling, different types of models are used, the dominating are statistical and data mining [5]. These include Exponential Smoothing Methods (ESM), Autoregressive Integrated Moving Averages (ARIMA) which is variants of Box-Jenkins methodology that follows three stages of model identification, estimation and validation [6] [7], Artificial Neural Network (ANN) variants Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), Long Short Term Memory (LSTM) . . . etc. These models can be applied separately or in combination for data prediction. Therefore, proper analysis of dependencies and similarities in network traffic with past observation will help in selecting the best accurate forecasting models, that is a close fit between the model and actual traffic trace [6] [7]. Accurate traffic prediction has many applications in network areas such as efficient utilization of network resources, design and capacity planning [1].

This study presents details of Ethio telecom broadband VSAT network data traffic analysis and modeling. That covers what behavior or patterns are exhibited when the data traffic is analyzed, which model is more suited to the network data, how models perform in testing process, and which forecasting model is the best from model comparison for the VSAT data traffic forecasting.

1.1 STATEMENT OF THE PROBLEM

Ethio Telecom is providing broadband service to its customers through fixed and wireless network. VSAT is one of the fixed wireless broadband network, which is delivering broadband services via satellite link. In addition, it is supporting backhaul service to UMTS mobile network. The services are available up to 5Mbps based on best effort shared bandwidth configuration. It is known that VSAT services are used as a last mile solution due to its delay and high service cost. However, Ethio Telecom has about 1,500 VSAT customers such as government and private companies. The service demand and traffic volume is increasing due to the introduction of high-data rate services, mobile users through cellular backhauling, and tariff reduction in VSAT services [2].

Currently, the VSAT networks are operating with limited satellite bandwidth, network resources, and processing capacity. Therefore, it is exposed to high congestion and poor QoS due to high bandwidth demand, bursting traffic, and round trip delay or latency. Besides, there is unbalanced IP traffic handling in gateway servers and no priority configuration based on traffic usage. Furthermore, there is a lack of trends to predict the traffic based on the network usage and demands in scientific ways. Usually, VSAT network optimization and expansion, which need high capital investment, are performed following the government policy and with market survey or analysis. As a result, logging and capturing the past network traffic utilization is important to understand the network traffic patterns and dynamics.

That is the main input for the planning and optimization and it has to be predicted as accurately as possible.

Therefore, it is needed to handle network traffic analysis and modeling for prediction including the insight for future planning and optimization, congestion control, network administration and traffic engineering. Accurate traffic models are also necessary for service providers like Ethio telecom to maintain the QoS properly, handle the increasing demand, allocate network resources appropriately, meet the Key Performance Indicator (KPI) settings and guarantee the subscribers' satisfaction. So that it will answer the questions how to understand the trends of traffic flows, how to improve the QoS, how to manage resource utilization, what will be the future traffic volume and how to design and optimize the network.

1.2 OBJECTIVES

1.2.1 *General Objectives*

The main objective of this study is to develop statistical and machine learning time series forecasting models for VSAT data traffic volume and choose the best one by evaluating the forecasting performances.

1.2.2 *Specific Objectives*

The specific aims of the study are as follows:

- To understand the operation and traffic flows of the brodband VSAT hub network.
- To study time series forecasting models for network data traffic dataset.

- To choose the appropriate software tool for analyzing, preprocessing, and modeling the collected time series dataset.
- To train multiple models, including statistical and machine learning models, for downloaded data traffic.
- To test models and assess their performance using accuracy metrics.
- To analyze and discuss the performance of the models, make recommendations, and identify future research areas.

1.3 SCOPE AND LIMITATIONS

In this study, we cover the VSAT network traffic analysis, modeling of univariate time-series dataset, and identify the best forecasting model for future network planning and optimization tasks to Ethio telecom. The research is restricted to data traffic volume, not covering the voice traffic. Besides, the collected data is limited to Ethio Telecom Digital Video Broadcasting-Satellite Second Generation (DVB-S2) VSAT hub network at Sululta station. The other networks are not considered due to lack of complete data.

1.4 CONTRIBUTIONS OF THE STUDY

The outcome of this study have significant implications for Ethio Telecom and similar research areas. The following are the research's contributions:

- Create awareness by investigating the most commonly used time series forecasting models both statistical and machine learning in more scientific ways, and as useful approach to forecast downloaded traffic volume and give contribution in the area.

- Suggest the most accurate model that improves the performance of forecasting to be used as input for Ethio Telecom in VSAT network planning and optimization.
- The research findings can be used as a reference or benchmark for future related works.

1.5 RELATED WORK

Several researchers studied time series modeling and forecasting of network data traffic. Some of them are reviewed and presented as follows.

The authors in [7] aimed to understand mobile data traffic growth in Ethio Telecom by applying time series forecasting techniques on 9 months dataset. They selected SARIMA model due to the observations of seasonality in the past data. The Autocorrelation Function (ACF), Partial ACF (PACF), and Hurst exponent estimation were used to identify the non-stationarity properties of the data. Akaike Information Criterion (AIC), and Bayessian Information Criterion (BIC) statistical tests were performed to estimate model parameters. In addition, R-statistical software applied as a tool. The performance of the models evaluated by Mean Absolute Percentage Error (MAPE) accuracy metric. Finally, SARIMA (2,0,1)x(0,1,1)7 was selected having better prediction accuracy. The authors were not considering other model types for performance comparison.

Based on [8], the author focused on the analysis on the time series models by taking online three datasets having different size (14 months internet traffic, 23 months electric load, and 16 years stock prices). In data preprocessing stage, Hampel filter used for outliers, and data split with 70/30%. Further 50 and 100% of the training data were used to find out the model performances. The study included models: Double Exponential Smoothing

(DES) and ARIMA, and the ANN variant Feed Forward Neural Network (FFNN). In addition, the MATLAB software implemented as a tool. The Mean Square Error (MSE), MSE[in%], and Root Mean Square Error (RMSE) accuracy metrics were applied to measure forecasting errors. Finally, the author obtained the result with ANN[14.27%], next ARIMA[17.82%], and the last DES[75.11%]. The models' performance improved when training data was set to 100% and big data used. Deep Neural Network (NN) models were not considered for big size and complex dataset.

The study in [9], discussed on different time series forecasting approach to optimize the capacity of the wireless backhaul network. They collected traffic dataset with sampled size of 1274 for point to point link utilization. They set 20% of the mean forecasted values for capacity optimization. Moreover, they selected Autoregressive (AR), SARIMA and MLP models, including RMSE and MAPE forecasting metrics. The R programming language used as a software tool. Finally, they obtained the lowest error by MLP for all links than AR and SARIMA models. They depicted that 90% of utilized links are < 70% of their planned capacity, and they showed 75-100% of the link utilization can be improved or wastage reduced.

According to the authors in [1], they were able to find an efficient predictor for time series network traffic by considering the metrics such as high accuracy, low complexity of computation, and low consumption of energy. They collected the datasets from three locations with 16, 20 and 4 real network traffic traces, and set 25/75% for data splitting. In addition, they included the models: Last Value (LV), Moving Average (MA), Autoregressive Moving Average (ARMA), DES, AR, ANN, and wavelet decomposition. They applied the metrics normalized MSE (NMSE) and the Energy-Error score that combines the energy consumption and accuracy. The MATLAB and GEMS software with Wattch power simulator were used as a software tool. Finally, based on the Energy-Error score metric minimizing both energy and prediction error, they found DES performed the best, ARMA was the second with minimal energy overhead, ANN was the third performed well but had high energy and computation overhead. They limited the training dataset to 25% only.

As per [10] the author performed modeling and forecasting of the mobile network data traffic. The author collected 15 months data traffic and used linear Guassian model for smoothing and estimating missing values. Furthermore, stationary test was done by using Augmented Dickey Fuller (ADF) and Kwiatkowski- Phillips-Schmidt-Shin (KPSS) statistical tests. The SARIMA and ANN variants: Auto-Regressive Neural Network (ARNN), Extreme Learning Machine (ELM), MLP, models were used including the hybrid of SARIMA and ANN models. The RMSE, Mean Absolute Error (MAE), MAPE, and Mean Absolute Square Error (MASE) were applied to evaluate the model performance. All the the process were done by R-programming language. Finally, the author obtained an improved performance by 12.4% when SARIMA and ELM were hybridized than the other single and combined models.

The author in [11] explained the improvement of the network traffic prediction by using the multiple-model fusing method. The author collected two traffic dataset with a size of 1000 from different locations. A wavelet transform was applied to decompose the data into approximate and detailed components. Then the ARIMA model was used to predict the approximate and Least Squares Support Vector Machine (LSSVM) model for detail components where Support Vector Machine (SVM) parameters were approximated by applying Improved Particle Swarm Optimization (IPSO) optimization. Besides, the Guass-Markov estimation algorithm was applied to form a combined ARIMA/IPSO-LSSVM model by fussing their predicted values. Accuracy metrics: RMSE, MAE, MAPE, Relative RMSE (RRMSE), Sum Squre Erorr (SSE), Theil Inequality Coefficient (TIC), the Index of Agreement (IA), R^2, and Reliability were used. Finally, ARIMA/IPSO-LSSVM model performed better reliability and valid prediction values in all metrics than individual models and other models previously performed.

In general, the researchers in the related works performed modeling and prediction on network data traffic using either a single or a combined model. For model forecasting processes, they used a variety of methods, techniques, and software tools. Furthermore, the impact of dataset sizes and its characteristics on model performance was demonstrated. This indicates that in order to recommend a model for forecasting time series data, it is necessary to use different of techniques and models ranging from simple to complex, such as classical, machine learning, and deep learning, in order to capture all characteristics of time-series network data traffic, make comparison and and determine the most accurate predictor.

1.6 METHODOLOGY

In this thesis, the following methodologies are followed to conduct the research:

- Related literature in univariate the time series forecasting models were reviewed.
- The VSAT network data traffic was collected from the Ethio Telecom network operation IP team and customer data from the VSAT team.
- The time series dataset was passed through preprocessing tasks to convert it into a convenient format using a Python programming language software.

- The data analysis and tests was performed to select the model order parameters, including the model training for both statistical and machine learning models (ESM, ARIMA, SARIMA, ANN)
- Evaluation of the model testing or forecasting performance handled by using time series accuracy metrics: RMSE, MAE, and MAPE.
- Comparison of models was done from the result of the model performances.
- The best forecasting model with the minimum errors was identified.

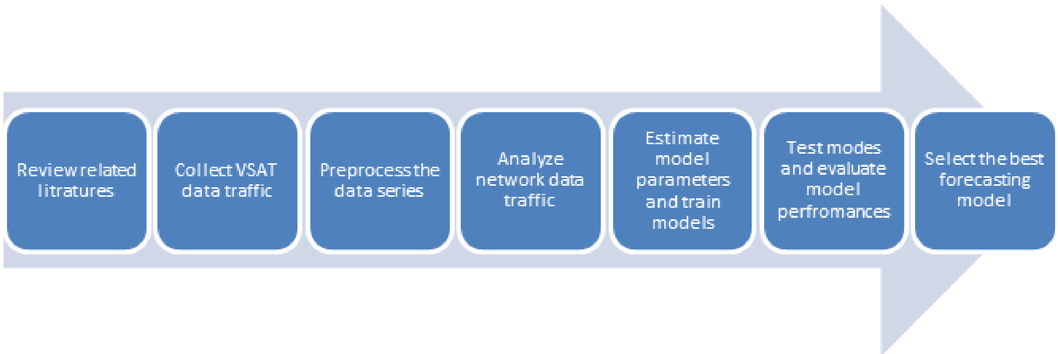

Figure 1.2: Methodology followed.

1.7 THESIS ORGANIZATION

The remaining of this thesis is organized in the following manner. Chapter two discusses the VSAT network system, architectures, and its applications. Chapter three deals with the time series data analysis and the principle of forecasting models which are applied in this thesis. In chapter four the details of the methodology followed in this thesis work are explained. Chapter five presents the result of the experiments and broadly discusses the findings. Finally, chapter six is all about the conclusion and recommendation including the future works.

2

OVERVIEW OF VSAT NETWORKS

The VSAT network is satellite-based interconnection and widely used in telecom industry due to its advantages of being easily installed and provide cost-efficient telecom services to residential, private, government and corporations. The VSAT network includes the hub station to integrate the VSAT to the IP core, the satellite system in space to re-transmit the incoming signal and the remote VSAT terminal equipment at the user side to deliver services [12] [13]. At present, the telecom VSAT network uses geostationary satellites orbiting the equator of the earth at the distance of 35,786 km. Due to such the long-distance, the radio frequency transmission power attenuated by 200dB, and having delay of 500msec for double hops [14] [12].

Figure 2.1 shows frequently used star and mesh topologies. Star networks are both unidirectional and bidirectional whereas mesh networks are bidirectional. Bidirectional star networks allow the transmission of information in both directions, but not directly from VSAT to VSAT, could be routed through VSAT hub, which requires double-hop connectivity and create propagation delay. In mesh VSAT networks, the remote terminals can transmit data directly to each other without passing through the central hub station. This allows to transmit high data rate and mixed traffic with a control station for signaling and resources allocation [15] [14].

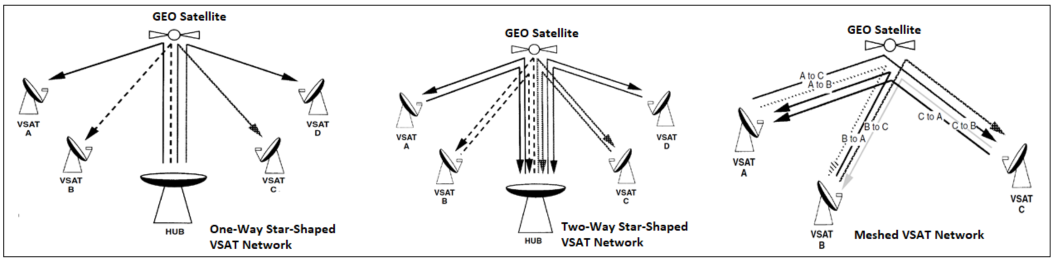

Figure 2.1: VSAT network topologies [12].

2.1 VSAT NETWORK IN ETHIO TELECOM

Currently, Ethio Telecom is operating three of VSAT hub networks at Sululta earth station. These networks are bidirectional star topology and configured with Digital Video Broadcasting-Satellite (DVB-S) standards (first, second and extension of second generation). Figure 2.2 shows the main components of the VSAT hub, and interconnection to the remote VSAT terminal for applications: internet, video or audio, Virtual Private Network (VPN) data and cellular backhaul services. The interface component contains the IP gateway servers to interconnect the hub and terrestrial data, managing point-to-point, broadcast, and multicast IP traffic to remote sites. The edge router or Aggregated Service Router (ASR) establishes a link to IP core network. In addition, it can log and store the network traffic data for analysis. The data for this study was captured from this router [16].

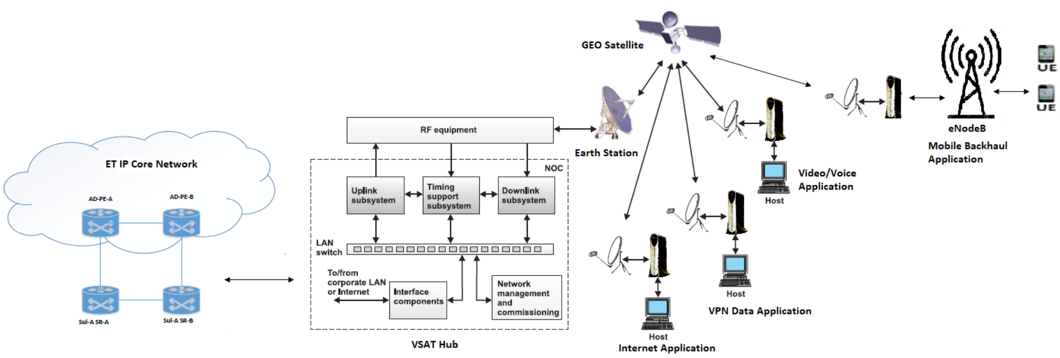

Figure 2.2: Ethio Telecom VSAT hub topology [16].

3

TIME SERIES FORECASTING MODELS

3.1 INTRODUCTION

A time series is an ordered observation taken in each time interval and generating series from a specific stochastic process. It is determined by equally spaced time stamped data-points. Time series data has characteristics of trend, which is long-term variations, and seasonal patterns of repeating cycles in time intervals [17] [18]. Usually, data are recorded in time series like hourly, daily, monthly, quarterly, and yearly. A univariate time series is a record of one variable, whereas multivariate time series is the record of more than variables [19]. Modeling in time series is to investigate the trend and seasonal patterns of the series and develop a univariate time series model to forecast the future values of the series [17].

3.1.1 *Time Series Analysis*

The essential characteristic of time series is that the current observation is highly dependent on the previous observation. Therefore, time series analysis is dealing with to study of practical technics to analyze the dependency of observations. The analysis needs the development of models for the time series data and forecasting of future values from current and past values [20]. In time series analysis, one begins from visualization by plotting the data. As a result, it is needed to decompose the series for analysis into a trend, seasonal and residual [21].

In addition, many techniques are studied to draw inferences from time-series data. Therefore, a hypothetical probability model is to be set up that represents the data. Then model parameters are estimated, the goodness of fit to the data to be checked and the fitted model is used to generate the series. Once the best model is obtained it is possible to apply it in areas of the field, like representing the trend of traffic volume demand, and others may use it to predict the future demand of the traffic volume [21].

3.2 STATISTICAL TIME SERIES MODELS

3.2.1 *Exponential Smoothing Models*

The ESM are considering only the moving average of the past observations. To forecast the next observation, they use the smoothed present values in the observed series [20]. The forecasting methods used in these models do not base on the statistical probabilities; rather the forecasting methods vary by simply changing the formulae in terms of level, trend, and seasonality smoothing parameters. As a result, three variants of ESM are available namely Single Exponential Smoothing (SES), Holt's Trend (DES) and Holt-Winter's seasonal (Triple Exponential Smoothing (TES)) [22].

3.2.1.1 *Single Exponential Smoothing Method*

The SES is considered as a baseline for comparison of other forecasting methods, and do not suitable for series having long-term trends and seasonal patterns. In other words, it is proper for series with constant mean that is changing slowly through time. SES is the simplest method that uses the geometric sum of past observation to forecast one step ahead value with a smoothing level parameter α ranging between 0 & 1, and α is estimated by minimizing the sum of squared forecasting errors during model training [22]. In other way, weights are given exponentially decreasing for different observations. Moreover, lower values of smoothing constant gives priority to past observation, and higher values to the most recent observations, but sensitive to any changes [23]. The general equation of SES is expressed as:

$$\hat{x}_N(1) = \alpha x_N + \alpha(1-\alpha)x_{N-1} + \alpha(1-\alpha)^2 x_{N-2} + ... \quad (3.1)$$

Equation 3.1 can be written as an estimation of the local level L_t at time t.

$$L_N = \alpha x_N + (1-\alpha)L_{N-1} \quad (3.2)$$

The h-step ahead forecast is:

$$\hat{x}_N(h) = L_N \quad (3.3)$$

3.2.1.2 *Holt's Linear Trend Method*

The Holt's linear trend method (DES) considers the local trend of the time series which implies proper forecasting series having trend patterns. This modifies the SES model by including one smoothing parameter β for the local linear trend. As a result, the local trend term T_t measuring of trend fluctuation per unit of time is included in SES and the modified equation is expressed as a generalized form given by L_N [22].

$$L_N = \alpha x_N + (1-\alpha)(L_{N-1} + T_{N-1}) \quad (3.4)$$

Updating the estimation of growth rate

$$L_N = \beta(L_N - L_{N-1}) + (1-\beta)T_{N-1} \quad (3.5)$$

Then, the h-steps-ahead forecast is given by

$$\hat{x}_N(h) = L_N + hT_N \quad (3.6)$$

Another parameter called trend damping ranges between 0 & 1 is to damp the estimated trend or its growth rate at time t, T_t to ϕT_t. Therefore, h-step ahead forecasting at time N with trend damping factor is expressed as [22]:

$$\hat{x}_N(h) = L_N + (\sum_{i=1}^{h} \phi^i)T_N \quad (3.7)$$

3.2.1.3 *Holt-Winters Seasonal Method*

The Holt-Winters seasonal method (TES) considers the trend and seasonal changes of the time series which is suitable for series having trend and seasonal behavior. Therefore, the exponential model is modified by introducing seasonal changes in the form of additive and multiplicative, the seasonal smoothing parameter γ, index of seasonal I_t with period time t, and number of seasonal period s. For additive seasonality, the indices are summed to zero, while for multiplicative seasonality, the indices are averaged to unity. Considering additive seasonality, the updated exponential smoothing equations with L_N level, T_N growth rate and I_N the seasonal index are denoted in equation 3.8, 3.9 and 3.10 respectively [22].

$$L_N = \alpha(x_N - I_{N-s}) + (1-\alpha)(L_{N-1} + T_{N-1}) \tag{3.8}$$

$$T_N = \beta(L_N - L_{N-1}) + (1-\beta)T_{N-1} \tag{3.9}$$

$$I_N = \gamma(x_N - L_N) + (1-\gamma)I_{N-s} \tag{3.10}$$

Then the h-step ahead forecast at time N for h=1,2,3,..., s is given by

$$\hat{x}_N(h) = L_N + hT_N + I_{N-s+h} \tag{3.11}$$

3.2.2 *Autoregressive and Moving Average Models*

3.2.2.1 *Autoregressive*

An AR, is a process of representing the current value at time t as a finite linear combination of previous values and the random noise. Assuming Z_t as the values of the process with equally spaced time *t-1, t-2*,... becomes Z_{t-1}, Z_{t-2},... And Z_t=Z_t - μ as deviation from the mean. Then AR process with order p, AR(P) is expressed as equation 3.12 which shows that the values of Z is regressed on its previous values Z_{t-1}, Z_{t-2}, ... etc [20].

$$Z_t = \sum_{j=1}^{p} \phi_j Z_{t-j} + a_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + ... + \phi_p Z_{t-p} + a_t \tag{3.12}$$

Where a_t is white noise $N(0, \sigma^2)$, ϕ_j is coefficients and p is positive integer.

An AR with order parameter p can be expressed in terms of backward shift operator B to define the lagged values of the process so $BZ_t = Z_{t-1}$, $B^2Z_t = Z_{t-2}, \ldots, B^dZ_t = Z_{t-d}$.

$$\phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p \tag{3.13}$$

Then AR(p) process can be expressed for t=1,...,n

$$\phi(B) Z_t = a_t \tag{3.14}$$

The $\phi(B)$ is referred to the polynomial of the AR process with degree P. The roots of $\phi(B) = 0$ defines stationary or not. For a stationary process, all roots must fall outside the unit circle and if there is a single unit root equals to 1, the process is non-stationary [20].

3.2.2.2 *Moving Average*

The MA is a process of representing observed values at time t as a finite number of q of previous random noise. Here the observed process depends linearly on the previous random noise. Therefore, the MA process Z_t with order q is defined as [20]:

$$Z_t = a_t + \sum_{j=1}^{q} \theta_j a_{t-j} = a_t - \theta_1 a_{t-1} + \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \tag{3.15}$$

Here the coefficients θ need not be unity nor positive. The MA(q) can be expressed in terms of backward shift operator B, for t=1,...,n.

$$\theta(B) = 1 + \sum_{j=1}^{q} \theta_j B^j = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q \tag{3.16}$$

Then MA(q) process can be expressed for t=1,...,n.

$$Z_t = \theta(B) a_t \tag{3.17}$$

3.2.2.3 *Autoregressive-Moving Average*

The ARMA is the unique representation of combining the two processes AR and MA to get better advantages and flexibility in fitting original time series. Hence, the ARMA process Z_t with p and q order is defined as [20]:

$$\begin{aligned} Z_t &= \sum_{j=1}^{p} \phi_j Z_{t-j} + \sum_{j=1}^{q} \theta_j a_{t-j} + a_t Z_t \\ Z_t &= \phi_1 Z_{t-1} + ... + phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - ... - \theta_q a_{t-q} \end{aligned} \tag{3.18}$$

The ARMA(p,q) can be expressed in terms of backward shift operator B.

$$\phi(B) Z_t = \theta(B) a_t, \text{for t=1,...,n} \tag{3.19}$$

Here ϕ(B) and θ(B) are the polynomials with a degree of p and q of ARMA process. For stationary time series, satisfactory representation can be established with AR, MA, or mixed ARMA models having order parameters p & q with a value of a maximum 2.

3.2.2.4 *Autoregressive Integrated Moving Average*

In several fields of area, the time series data manifest non-stationary characteristics and they do not change around fixed mean, however, they may produce the same kind of behavior through time. Therefore, the process to constitute a non-stationary series is given by ARIMA. The ARIMA process applies differencing on time series several times to reduce to ARMA process [20]. Once the time series is changed to stationary then the low order ARMA(p,q) model will be applied. The actual time series data will be processed by using ARIMA with order parameters p, d, and q [21]. Similarly, if the roots of φ(B) = 0 are inside the unit circle then the ARIMA process is non-stationary; otherwise, the ARIMA process stationary. Therefore, if there are d unit roots and others outside the unit circle, then, the operator φ (B) can be defined as [20].

$$\varphi(B) = \phi(B)(1 - B)^d \tag{3.20}$$

Where $\phi(B)$ is a stationary AR operator and the model representing the homogeneous behavior is expressed as:

$$\varphi(B)Z_t = \phi(B)(1-B)^d Z_t = \phi(B)\nabla^d Z_t = \theta(B)a_t \tag{3.21}$$

where

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$$

Therefore, the series having homogeneous non-stationary behavior can be transformed to stationary by applying the d^{th} difference on the ARIMA process, where the d value is at most 2 and with d=0 the series is stationary.

To define the forecasting of the ARIMA models, we start by letting the observed series process Z_t meets the difference equation [21]:

$$(1-B)^d X_t = Y_t, \ t = 1, 2, ..., \tag{3.22}$$

Where Y_t is ARMA(p, q) process and X_t is the random vector $X_{1\text{-}d}, \ldots, X_0$ which is uncorrelated with Y_t.

Then $P_n X_{n+h}$ is h-step prediction and expressed as equation 3.23:

$$P_n X_{n+h} = \sum_{j=1}^{p+d} \phi_j^* P_n X_{n+h-j} + \sum_{j=h}^{q} \theta_{n+h-1,j}(X_{n+h-j} - \hat{X}_{n+h-j}) \tag{3.23}$$

where

$$\phi^*(Z) = (1-Z)^d \phi(Z) = 1 - \phi_1^* Z - ... - \phi_{p+d}^* Z^{p+d}$$

P_n is used as an indicator of a linear predictor of the observation up to time n, with t=n+h and h is time-steps.

3.2.2.5 *Seasonal ARIMA*

The SARIMA process is an extension of the ARIMA process that considers seasonal behavior and shows dependency on the periodic characteristic repeating with a certain time of period s. For time series, there may be more than one seasonal period where similar patterns occur at s. Multiplicative time series models are used to handle times series, which have dependencies on the seasons, and between adjacent values. The current observation Z_t at time t can be linked with the previous observation by a simplified form of model with seasonal backshift B^s, backward difference operators ∇_s, and seasonal period s. This is expressed as equation 3.24 and meets the stationarity condition of the time series[20].

$$\Phi(B^s)\nabla_s^D Z_t = \Theta(B^s)\alpha_t \tag{3.24}$$

Where ∇_s = 1 - B^s, $\Phi(B^s)$ and $\Theta(B^s)$ are polynomials in B^s of degrees P and Q respectively which are order parameters for seasonal behavior.

In this model, there would be a correlation in the error components α_t to α_{t-1} and to α_{t-2} ... etc. So that another model is added to represent their relationship.

$$\phi(B)\nabla^d \alpha_t = \theta(B)a_t \tag{3.25}$$

Where ∇^d = 1 - B, $\phi(B)$ and $\theta(B)$ are polynomials in B of degrees p and q respectively which are order parameters for non-seasonal behavior.

Finally, we will get a general multiplicative model by substituting equation 3.25 in 3.24.

$$\phi_p(B)\Phi_p(B^s)\nabla^d\nabla_s^D Z_t = \theta_q(B)\Theta_Q(B^s)a_t \tag{3.26}$$

Where s, p, P, q, Q, d, and D are the order parameters for the seasonal ARIMA model. So the multiplicative process with its all orders will be denoted by SARIMA(p,d,q)x(P,D,Q)[s] [20].

To forecast SARIMA models, similar process will be followed as to forecasting ARIMA models shown in section 3.2.2.4. The best linear predictors $P_n X_{n+h}$ of X_{n+h} is expressed as:

$$P_n X_{n+h} = P_n Y_{n+h} + \sum_{j=1}^{d+Ds} a_j P_n X_{n+h-j} \tag{3.27}$$

Where $P_n Y_{n+h}$ is linear predictor of Y_t which is the ARMA process in terms of 1, Y_1,...,Y_n. Then, the $P_n X_{n+h}$ can be calculated recursively for $h \geqslant 1$ [21].

3.3 MACHINE LEARNING TIME SERIES MODELS

3.3.1 *Artificial Neural Networks*

The ANN provides to model the non-linear series data. It does not need any assumption on the statistical nature of the series data before training the model. The structure of ANN is developed by imitating the human brain system. The NN is a system in which a set of inputs and outputs are connected in a non-linear manner. So, the NN architecture defines the connection between input and output layers, the number of layers, and the number of neurons grouped in each layer. In time series, the inputs are sequential lagged values and the output are the forecasted values [22] [24] [10]. The ANN is represented by simplified model in Figure 3.1 comprised of connecting links having weights to measure the strength of connection and produced weighted input signal, summing component outputs the linear sum of weighted input signal and bias defined by each neuron, and the activation function transforms into a non-linear output signal [8].

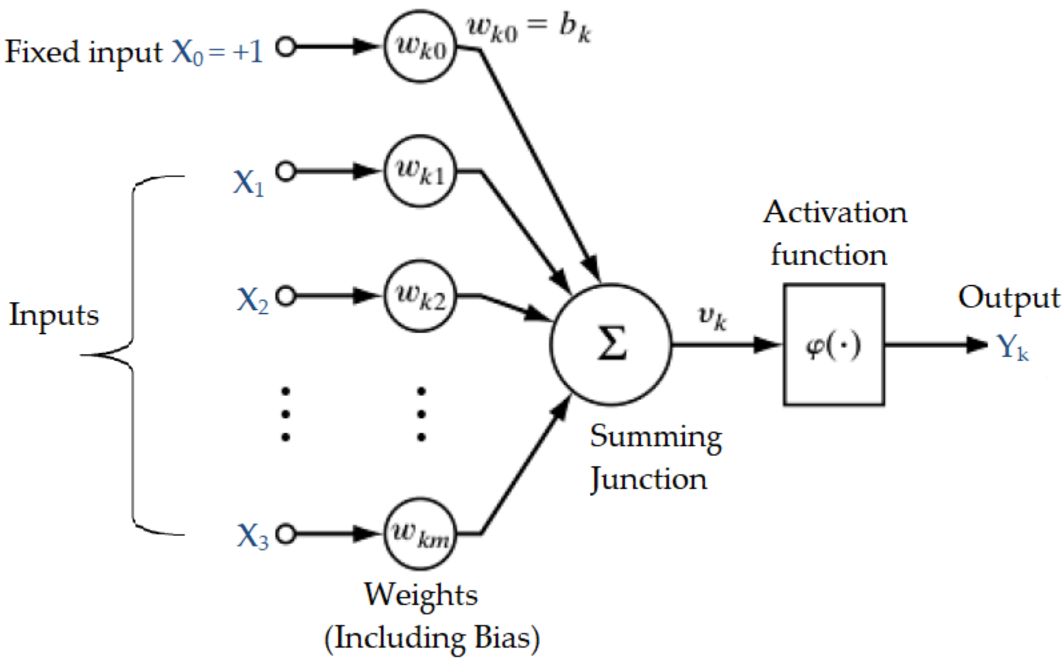

Figure 3.1: ANN simplified model [8].

The ANN model is expressed by equation 3.28:

$$v_k = \sum_{i=0}^{m} w_{kj}.x(j) \tag{3.28}$$

Where x_0=1 and x_1, x_2, ... x_m are the input signals, w_{k1},...,w_{km} are the weights of neuron k, b_k is the bias and v_k is the linear sum of weighted signal. The output signal of the neuron y_k is expressed as equation 3.29 where φ(.) represents the activation function.

$$y_k = \varphi(v_k) \tag{3.29}$$

The nodes in each layer will perform the process of the linear sum of the weighted inputs and then transform them into non-linear by using different activation functions. The sigmoid, hyperbolic tangent, rectified linear unit (ReLU), and leaky rectified linear unit (LReLU) activation function are usually used [10] [8] [24].

3.3.2 *Multilayer Perceptron*

The MLP is basic and a commonly used type of NN. It can be a type of single hidden layer (single perceptron) or more in a group of networks that form a feedforward NN architecture where they direct signals from input to output in one direction and there are no feedback loops inside the network. Each neuron in one layer is directly connected to the neuron in the next layer forming mesh connections shown in Figure 3.2 [8] [19] [22] [24].

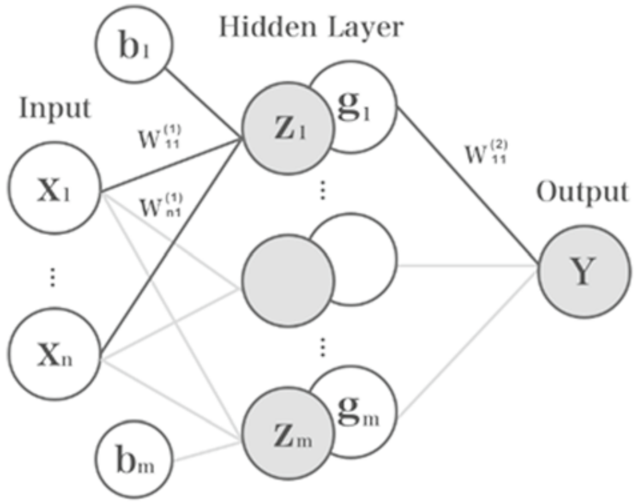

Figure 3.2: MLP feedforward network diagram [24].

From Figure 3.2 the output of the neural network is expressed as:

$$Z = W^T X + B \tag{3.30}$$

where Z is a vector after a linear combiner, W^T is a transpose weight matrix, the input X $=X_1,\ldots,X_n$ and a bias vector for each hidden node B = $b_1,\ldots, b_n$. Then the output of neural network Y is given by:

$$Y = W_0^T G + B_0 \tag{3.31}$$

Where W_o is the weights of the output and b_o is biases for each output, G is An activation function such as the sigmoid function $G = \mathtt{Sigmoid}(Z)$ and $W_o = [W_{ij}]_{m\ x\ 1}$.

3.3.3 *Recurrent Neural Networks*

The RNN is similar to deep-learning neural networks where there are a group of networks and a feedback loop inside the network. RNN uses current and previous observations to build the sequential or temporal models by learning the temporal relationship of the data and the target label. However, RNN is suffering from the vanishing gradients problem so it is hard to train the model for long-term temporal correlations. The Figure 3.3 shows the RNN network architecture [8] [19] [24]:

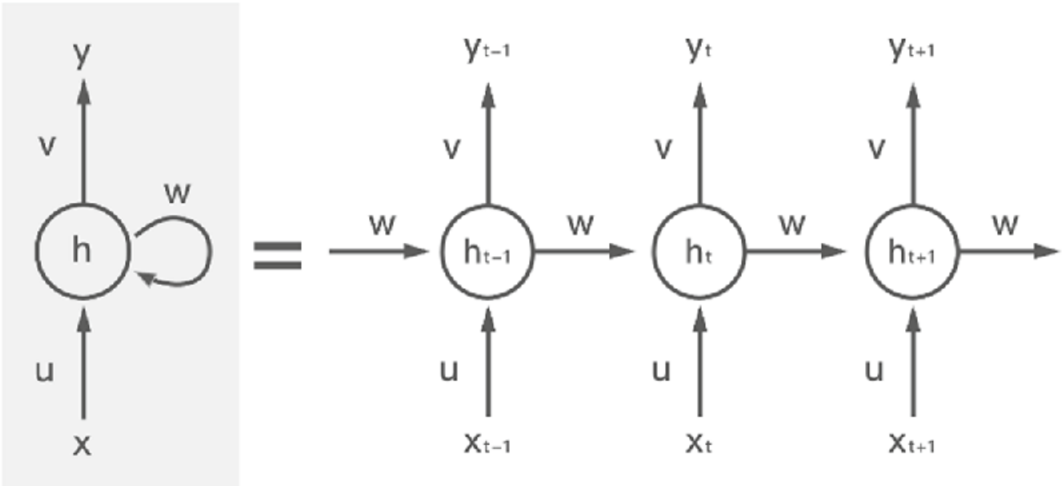

Figure 3.3: RNN model architecture [24].

From Figure 3.3 x_t is the input at time step t, x_{t-1} is the previous step and x_{t+1} is the next step t. u and v are weights, and y_t is the target label at time t. A vector $x(t)$ is represented by linking a vector at time t and h at time t-1. The j^{th} hidden layer state $h_j(t)$ and the k_{th} target label. All are expressed as equation 3.32 and $f(.)$ is the a transfer function such as the sigmoid.

$$
\begin{aligned}
x(t) &= w(t) + h(t-1) \\
h_j(t) &= f(\sum_i x_i(t) \times u_{ji}) \\
y_k(t) &= f(\sum_j h_j(t) \times v_{kj})
\end{aligned}
\tag{3.32}
$$

3.3.4 *Long Short Term Memory*

The LSTM is one of the deep-learning NN and an extension of the RNN that includes unique elements of memory cells and gates inside the network to solve the vanishing gradient problem. LSTM is able to learn long-term dependencies. Figure 3.4 depicts the LSTM network architecture [24] [19].

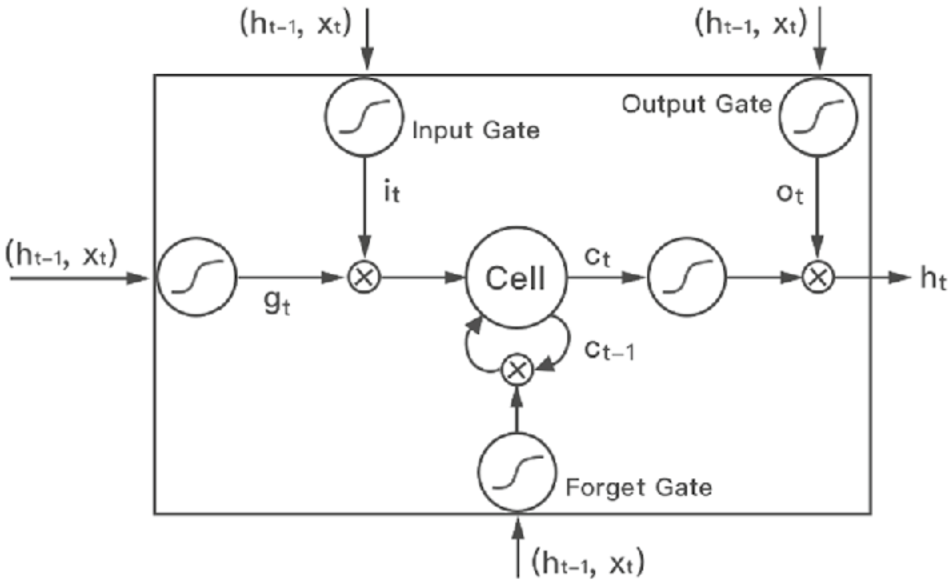

Figure 3.4: LSTM model architecture [24].

The LSTM network elements in Figure 3.4 are mathematically expressed as equantion 3.33.

$$\begin{aligned}
i_t &= \mathrm{Sigmoid}(W_i[h_{t-1}, x_t] + b_i) \\
f_t &= \mathrm{Sigmoid}(W_f[h_{t-1}, x_t] + b_f) \\
o_t &= \mathrm{Sigmoid}(W_o[h_{t-1}, x_t] + b_0) \\
g_t &= \mathrm{Tanh}(W_g[h_{t-1}, x_t] + b_g) \\
c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\
h_t &= o_t \odot \mathrm{Tanh}(c_t)
\end{aligned} \tag{3.33}$$

Where the input gate, forget gate, output gate, and input node are i_t, f_t, o_t, g_t and W_i, W_f, W_o, W_g are weight vectors for the corresponding inputs respectively. The symbol $\odot$ indicates the element-wise product. The components of the LSTM network are explained as follows [24].

- **Input node (g_t)**: updates x_t of the current state of the input layer and $h_{t\text{-}1}$ of the previous status of the hidden layer at current time step t.
- **Input gate**: determines if the input node g_t value to go into the new cell. If the value of the input gate is near zero, the gate will be closed so that g_t is prohibited from entering.
- **Forget gate**: helps the network either to keep or erase the previous state $h_{t\text{-}1}$ value in a memory cell. The memory cell will be erased if the forget gate value is zero.
- **Output gate**: determines the output h_t from the cell state.

Forecasting Neural Network Models: forecasting on the output is performed based on the observation of the lagged past values in sequence as input with hidden layers of neurons. In time series, before the output the series must be rescaled to ranges for Sigmoid(0,1), otherwise the output will be in wrong order of magnitude or use identity activation function. The prediction that provides an h-steps ahead for simple NN with one hidden layer of H neurons, the past lagged observations $x_{t\text{-}j_1}, \ldots, x_{t\text{-}j_k}$ as input and the output x_t could be expressed as equation 3.34 [22]

$$\hat{x}_t = \phi_o(w_{co} + \sum_{h=1}^{H} w_{ho}\phi_h(w_{ch} + \sum_{i=1}^{k} w_{ih}x_{t-j_i})) \tag{3.34}$$

Where w_{ch} is weight between the constant input and the hidden neurons h, w_{co} weight between constant input and the output, w_{ih} weights for the other connections between the inputs and the hidden neurons, and w_{ho} the weights between the neurons and the output. Moreover, ϕ_h and ϕ_o are the activation functions used at the hidden layer and the output, respectively.

4

EXPERIMENTAL DESIGN

4.1 SYSTEM MODEL

The system model for this thesis shown in Figure 4.1 starts from the dataset collection and completion of the data preprocessing. Then, the dataset is partitioned into 80% train set and 20% test set for the statistical (ESM, ARIMA, SARIMA) and machine learning ANN (MLP, RNN, LSTM) models.

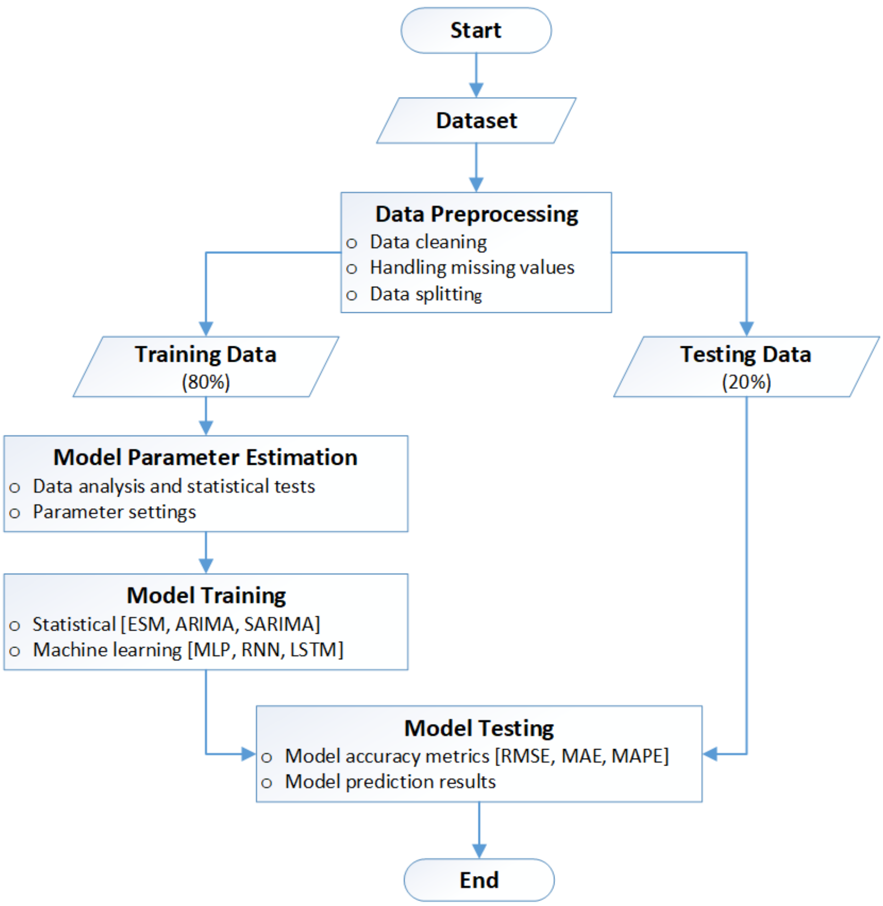

Figure 4.1: System model.

The 80% train data is used for analysis, model training to get the efficient model parameters. The remaining 20% test data is set as out of sample data and is used to evaluate the prediction of the models. Finally, the performance of the models is compared by using different metrics.

4.2 SOFTWARE

In this thesis, we used an open-source software called Python programming language in Jupyter notebook environment with Scikit-learn library to compute data preprocessing, data analysis, model training, and forecasting for statistical models. In addition, for NN deep learning models, Keras library and Tensorflow as backend are used in the Jupyter notebook [24] [25].

4.3 DATASET

In this thesis, the VSAT network traffic dataset was obtained from Ethio Telecom national operation center team. The data traffic was captured from edge router (ASR) located at the satellite station, which contains the uplink and downlink traffic volume per VSAT network. The span of the network data traffic is 12 months from 01-Mar-2020 to 28-Feb-2021 with an hourly-based measurement saved in Comma Separated Values (CSV) file format. We used traffic data for the DVB-S2 VSAT network and considered its downlink traffic volume measured in Mbytes as an attribute with 8760 data points.

4.4 DATA PREPROCESSING

The raw dataset contains network data traffic recorded with an interval of an hour and saved in CSV file format for 12 months. In data preprocessing, the raw time series dataset has to pass through data transformation processes before being applied to modeling and forecasting. Therefore, we handled data cleaning that includes removing unwanted suffixes and non-numerical strings in each feature one-by-one. Then, for each date and time, the corresponding recorded data points of the downlink traffic volume were saved as a single new dataset by concatenating all 12 monthly data in one CSV file. Finally, the yearly univariate time series dataset is used for traffic analysis, modeling, and forecasting processes. The sample data from yearly dataset is shown in the Table 4.1.

Table 4.1: Monthly sample time series data.

Date_Time	**Traffic_DL_MB**
4/1/2020 0:00	2674.00
4/1/2020 1:00	1890.00
4/1/2020 2:00	1740.00
4/1/2020 3:00	1597.00
4/1/2020 4:00	1209.00
–	–
4/30/2021 19:00	6222.00
4/30/2021 20:00	6366.00
4/30/2021 21:00	7081.00
4/30/2021 22:00	5281.00
4/30/2021 23:00	2224.00

720 rows x 1 columns

4.4.1 *Missing Values*

We identified the missing values by counting empty or Not a Number (NaN) values, and by visualizing the plots of the data series. Numerically

177 NaN values were counted which accounted for 2.02% of the actual data size. In addition, from the graphical representation of the data in Figure 4.2 (a), we can observe consecutive zero values lasting for one week from 2020-07-14 19:00:00 to 2020-07-22 11:00:00, counted 185 which were happened due to a service outage in the station. Therefore, we considered this interrupted outage time as missing values so we replaced them by NaN. As a result, the total missing value is increasing to 362, 4.13% and shown in Figure 4.2 (b).

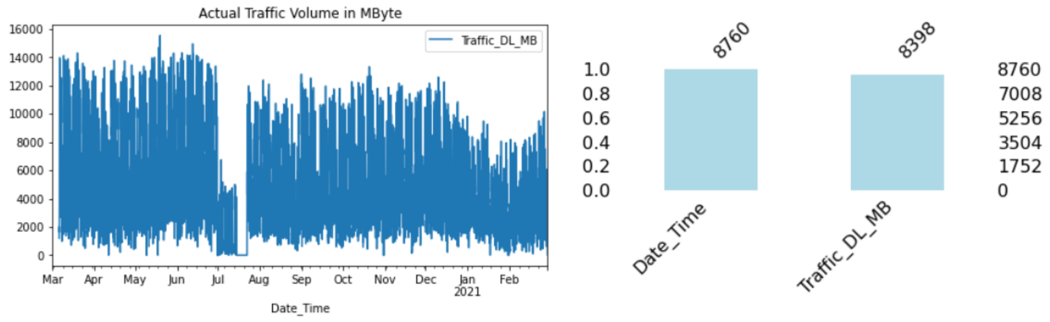

Figure 4.2: (a) Actual plots of traffic volume, (b) Bar plot for missing values.

Finally, as the total missing values 4.13% are less than 20%, it is recommended to replace instead of discarding them [18]. Therefore, we estimated the missing values by Linear Regression (LR) interpolation method, commonly used for time series data to predict the missing values [24]. Then the NaN values became null shown in Figure 4.3 (b). So that the time series data became consistent and complete for the analysis process.

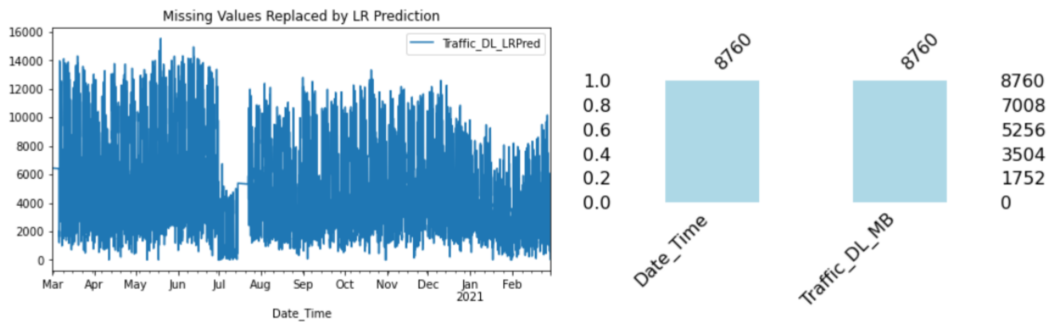

Figure 4.3: (a) Plot after NaN values estimated by LR, (b) Zero NaN values.

4.4.2 *Outliers Detection*

Outliers in time series data are values that are far apart from other observations. Outliers in the dataset may represent correctly or incorrectly recorded data. However, their presence will reduce the performance of the model forecasting, so identifying outliers is critical [8]. The interquartile range (IQR) method is a popular method for detecting and removing outliers in time series data. As illustrated in Figure 4.4 the IQR is the difference between the first (25^{th} percentiles) and third (75^{th} percentiles) quartiles of a set of data, with outliers observed far from the main concentration of the data. Outliers are defined as any value $1.5 * \text{IQRs}$ less than the first quartile or $1.5 * \text{IQRs}$ greater than the third quartile using a whisker width of 1.5 [27] [28] [29].

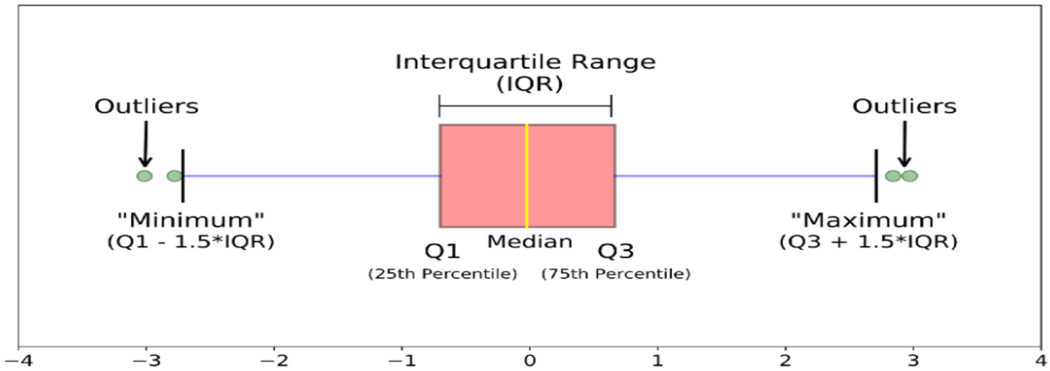

Figure 4.4: Components of Box-and-Whisker plot [28].

In this thesis, we used the IQR method to detect outliers and replace them with the upper limit value rather than removing them in order to keep the information while not reducing the dataset size. As a result, the box-and-whisker plots in Figure 4.5 show that outliers were detected and replaced.

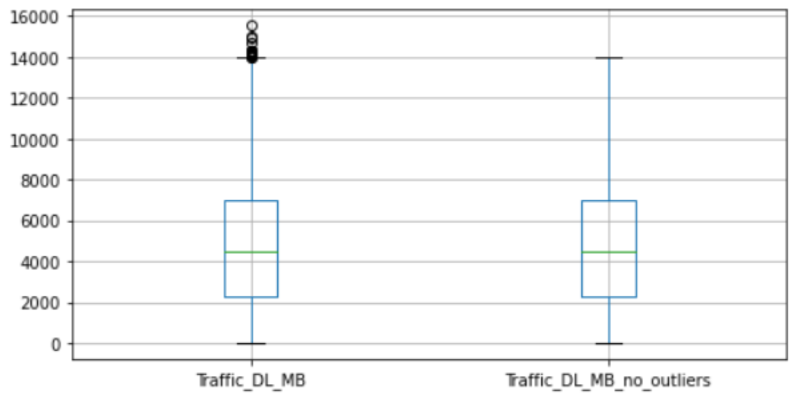

Figure 4.5: Filtered outliers with IQR method.

4.4.3 *Data Scaling*

Normalization is rescaling the data values into some specific upper and lower limits (-1, 1). It is also required, and essential in time series models to normalize the data having input variables with different scales [8]. In NN it improves the stability and performance by increasing the speed of training process and convergence [30]. In this thesis, we applied the min-max scaling method in the range of (1, 2). The general equation for rescaling the data into an arbitrary range of (a, b) is expressed as equation 4.1 [31]:

$$x^{'} = a + \frac{(x - min(x))(b - a)}{max(x) - min(x)} \tag{4.1}$$

Where $min(x)$ and $max(x)$ corresponds to the minimum and maximum values of the time series x(t). a and b are the required ranges.

4.4.4 *Data Partitioning*

In comparative forecasting analysis, it is practical to divide the dataset into two parts of the training set and the testing set. The training set is applied to model training. The testing set is reserved to handle out-of-sample forecasting and compared with the actual observation [22]. When dealing with a time series dataset, the current data points depend on the previously observed data so that the dataset should not be split randomly [32]. In this thesis, we partitioned the dataset with 80% of the first part for training data and the remaining 20% for testing data. The Figure 4.6 illustrates the partitioned dataset by applying an 80:20 ratio and the respective size of the train and test datasets.

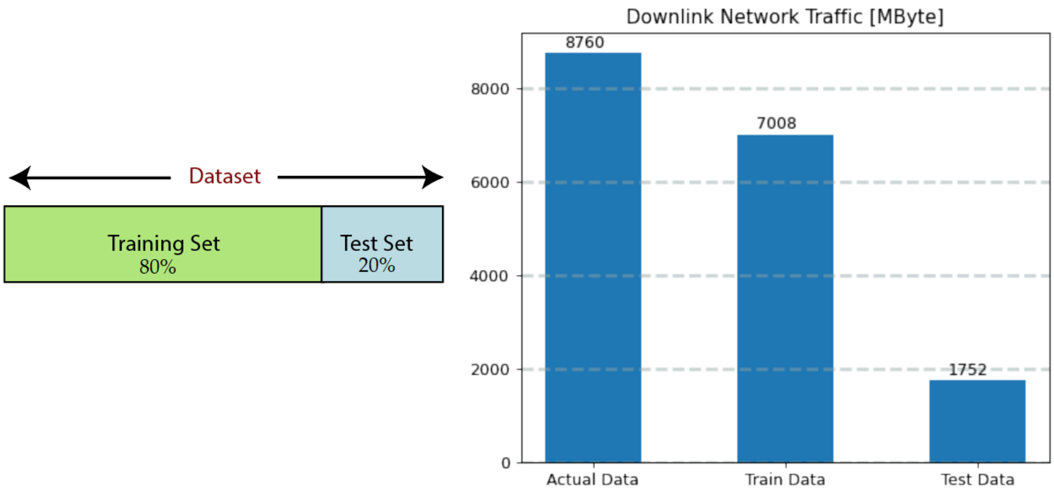

Figure 4.6: Dataset splitting into train and test sets.

4.5 DATA ANALYSIS

4.5.1 *Visualization*

In time series analysis, one begins from visualization by plotting the data. From the output, there are possibilities to observe or inspect the lack of continuity of series, unexpected change of levels, and outliers [21]. As a result, a visual analysis is performed by plotting the data as shown in Figure 4.7, where there are irregular high spikes and repeating patterns observed on the daily, weekly and monthly data traffic. In addition, there is high traffic demand during working hours 10:00 to 12:00 and 14:00 to 17:00 on weekdays, whereas it is decreasing on weekends.

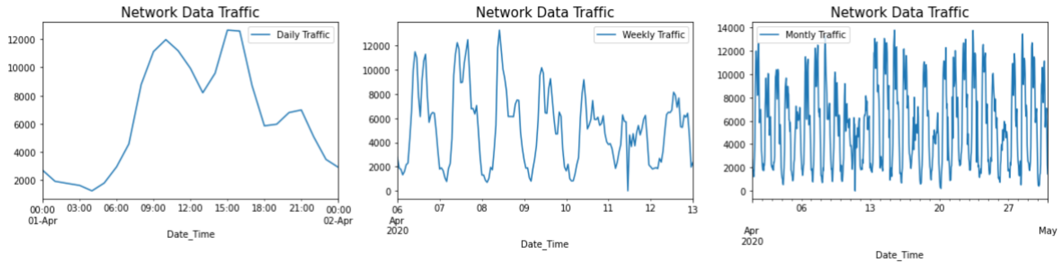

Figure 4.7: Network data traffic trace for daily, weekly and monthly.

4.5.2 *Decomposition*

There are different methods to estimate and decompose or eliminate the components of time series for analysis. The common estimation methods are smoothing with moving average filtering, exponential smoothing, smoothing by elimination of high frequencies, and polynomial fitting. Besides, the differencing method is used for elimination of components [21]. In general, there are two types of models in time series. The multiplicative model not all components are essentially independent. However, in the additive model, all components are independent. The multiplicative and additive models are expressed in equations 4.2 and 4.3 respectively [19].

$$X_t = T(t) * S(t) * R(t) \tag{4.2}$$

$$X_t = T(t) + S(t) + R(t) \tag{4.3}$$

Where T(t), S(t), and R(t) are trends, seasonal and residual respectively.

In this thesis, the additive and multiplicative decomposition of dataset is done. From the data decomposition, there are seasonality patterns with repeating cycles and fluctuating trends. The Figure 4.8 shows the multiplicative decomposition of partial dataset of the time series due to the dependency of trend and seasonal components.

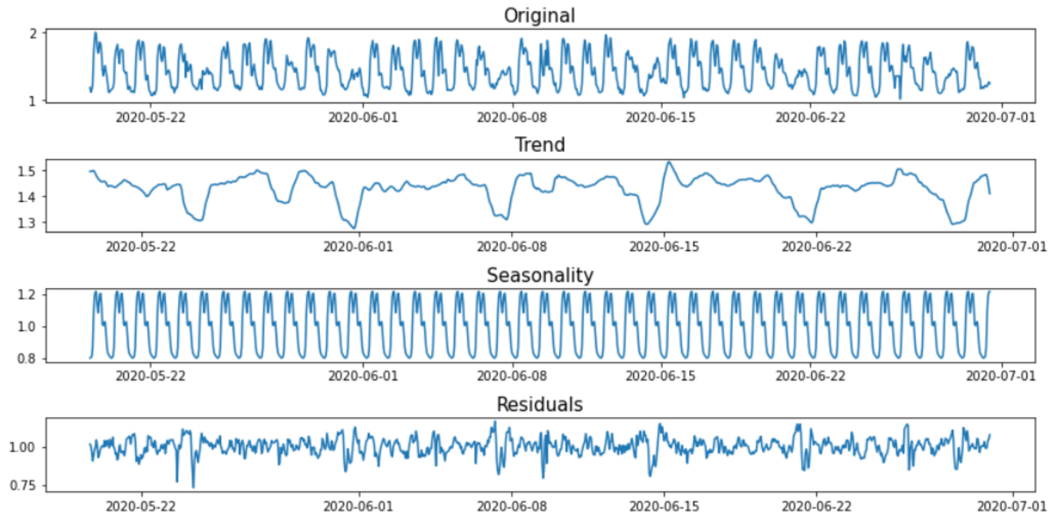

Figure 4.8: Time series decomposition, Multiplicative.

4.6 MODEL SELECTION AND PARAMETER ESTIMATION

The model parameters are identified by estimation. Therefore, it is essential to use statistically "inefficient" methods or techniques for a good decision like mathematical estimation, graphical observations, and so on [20]. Once the model is found by estimating the optimal parameters then the future values can be forecasted [8].

4.6.1 *Exponential Smoothing Orders*

Time series forecasting in ESM, the weights are given in exponentially decaying to the past observations. So last observations are considered to predict the future. The 3 variants of the ESM differ by their smoothing order parameters according to their characteristics. SES uses α level smoothing constant, DES adds β trend component and TES adds γ seasonal component. Besides DES and TES uses optional trend damping factor ϕ. All have values ranging (0, 1). The forecasting of the exponential models can be done either by finding the best values of smoothing parameters or by the auto-optimization method finding the minimum error of the smoothing curve [8] [36].

In this thesis, the ESM order parameters α, β and γ were selected with range of (0, 1). Trend damping factor ϕ=0.8 used in DES and TES. The Multiplicative seasaonal option used for TES as per the decomposition shown in section 4.5.2. Then the best one with minimum RMSE and MAE was selected in the model building and forecasting processes.

4.6.2 *Stationarity Test*

Knowing a unit root problem when the polynomial of the statistical time series model having unit root (near to unit circle) helps to decide differencing to make time series stationary before the data is applied to model training. Dickey and Fuller developed a systematic approach to identify the presence of unit root series based on the conditional least-square estimator for AR process and its "t-statistic". For the time series variable X_t follows AR process and for differenced series, the model can be written as [21] [20]:

$$\nabla X_t = \phi_0^* + \phi_1^* X_{t-1} + ... + \phi_p^* X_{t-p+1} + Z_t \tag{4.4}$$

Where $\phi_0^* = \mu(1 - \phi_1 - ... - \phi_p), \phi_1^* = \sum_{i=1}^{p} \phi_i - 1, \phi_j^* = -\sum_{j=1}^{p} \phi_i, \text{and}, j = 2, ..., p$

The limit distribution as $n \longrightarrow \infty$ of t-ratio derived by Dickey and Fuller the same as the test statistics.

$$\hat{\tau}_\mu := \hat{\phi}_1^* / \widehat{SE}\left(\hat{\phi}_1^*\right) \tag{4.5}$$

Where $\widehat{SE}$ is the estimated standard error of $\hat{\phi}_1^*$.

Then a test of null hypothesis Ho: can be established with the unit root assumption ϕ_0^* = 0. Having the t-ratio $\hat{\tau}_\mu$ 0.01, 0.05, and 0.10 quantiles of its limit distribution is to be -3.43, -2.86, and -2.57, respectively. By setting the level or quantiles let say 0.05, a unit root's the null hypothesis is rejected if $\hat{\tau}_\mu$ < -2.86 [21]. The KPSS stationary test is similar to the ADF except that the approach is an alternative way of testing the null hypothesis. That is, the stationary time series has a p-value greater than 0.05 [7].

In this thesis, we applied the ADF and KPSS tests to identify the time series whether weak or strong stationary, and to find the length of differencing. The time series is stationay for ADF and non-stationary for KPSS test. Therefore, the time series data is difference-stationary and it becomes strictly stationary after first-order differencing d=1 is applied, both tests in Figure 4.9 show the time series become stationary [37]. In addition, seasonal differencing D=1 is needed to make the series stationary.

```
Results of Dickey-Fuller Test:
Null Hypothesis: Unit Root Present
Test Statistic < Critical Value => Reject Null
P-Value <= Alpha(.05) => Reject Null

Test Statistic                   -21.446321
p-value                            0.000000
#Lags Used                        34.000000
Number of Observations Used     8724.000000
Critical Value (1%)               -3.431100
Critical Value (5%)               -2.861871
Critical Value (10%)              -2.566946
dtype: float64
```

```
Results of KPSS Test:
Null Hypothesis: Data is Stationary/Trend Stationary
Test Statistic > Critical Value => Reject Null
P-Value <= Alpha(.05) => Reject Null

Test Statistic            0.001415
p-value                   0.100000
Lags Used                32.000000
Critical Value (10%)      0.347000
Critical Value (5%)       0.463000
Critical Value (2.5%)     0.574000
Critical Value (1%)       0.739000
dtype: float64
```

Figure 4.9: ADF and KPSS test results for differenced data.

4.6.3 *Autocorrelation Functions*

In time series, ACF and PACF are used as a tool to identify the degree of dependence and to select the models. ACF tells the correlation between the time series data with its lag values. The ACF is expressed as a ratio of covariance of the stationary stochastic time series X_t at lag h and at 0 lag (a constant variance σ^2 for all t) [21].

$$\rho_X(h) = \frac{Cov(X_t, X_{t+h})}{\sqrt{Var(X_t)Var(X_{t+h})}} = \frac{\gamma_X(h)}{\gamma_X(0)} \tag{4.6}$$

The confidence interval of ACF with 95% should fall within the interval of $\pm 1.96/\sqrt{n}$, where 1.96 is the 0.975 quantiles of the standard normal distribution and n is number of observed samples [21].

The PACF can be used if we want to identify the correlation between the stationary processes X_t and X_{t+h} after eliminating the mutual linear dependence that occurred in between variables $X_{t+1}, \ldots, X_{t+h-1}$ [17].

$$P_h = \frac{Cov\left[(X_t, \hat{X}_t), (X_{t+h}, \hat{X}_{t+h})\right]}{\sqrt{Var(X_t)}\sqrt{Var(X_{t+h})}} \tag{4.7}$$

Where $\hat{X}_t$ and $\hat{X}_{t+h}$ are the best linear estimate of X_t and X_{t+h}, respectively.

From the plots of ACF and PACF, general rule of thumb can be applied to determine the properties and to select the low order parameters for autoregressive models based on Table 4.2. For example the AR model will be selected if the ACF graph trails off and the PACF has high spikes (strong lags), then the order of AR is equal to the number of q significant values in the PACF plot. Moreover, the ARIMA model will be selected if the series needs differencing to be stationary [38] [18].

Table 4.2: ACF and PACF model selection guidelines [18].

Models	**AR(p)**	**MA(q)**	**ARMA(p,q)**	**SAR(P)**	**SMA(Q)**	**SARMA(P,Q)**
ACF	Tails off	Cuts off at lag q	Tails off	Tails off at lag P*s	Cuts off at lag Q*s	Tails off at lag P*s
PACF	Cuts off at lag P	Tails off	Tails off	Cuts off at lag P*s	Tails off at lag Q*s	Tails off at lag Q*s

In this study, we used ACF and PACF plots including the guidance put in the Table4.2 to estimate the parameters of the autoregressive models. The Figure 4.10 is the ACF plots of actual and two-hourly resampled data, and shows repeated patterns for both data so that it indicates to handle trend and seasonal differencing on the data. In addition, Figure 4.10 (b) shows that there are spikes at every 12 lags but no other significant spikes in ACF, so that it determines the seasonal period s=12.

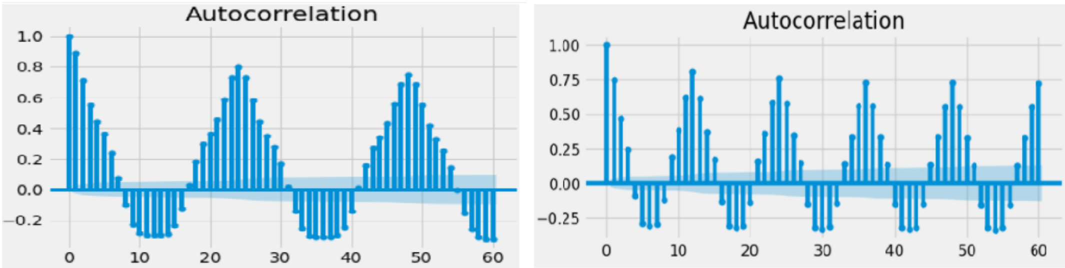

Figure 4.10: ACF plots (a) actual data, (b) two-hourly resampled data.

In Figure 4.11 and 4.12, we showed the plots of ACF and PACF for the differenced and eliminated seasonal component from the data to estimate the possible number of order parameters of ARIMA(p,d,q) and SARIMA(P,D,Q)[s] models. Note that we have used two-hourly re-sampled dataset for SARIMA model due to its high computational time and need of resources. As a result, we selected the models with optimal parameters ARIMA(0-2, 1, 0-2) and SARIMA(0-2, 1, 0-1)[12] for training and testing the series.

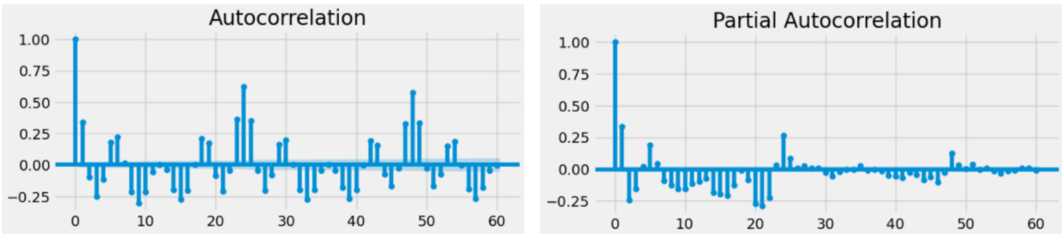

Figure 4.11: ACF and PACF plots for differenced actual data.

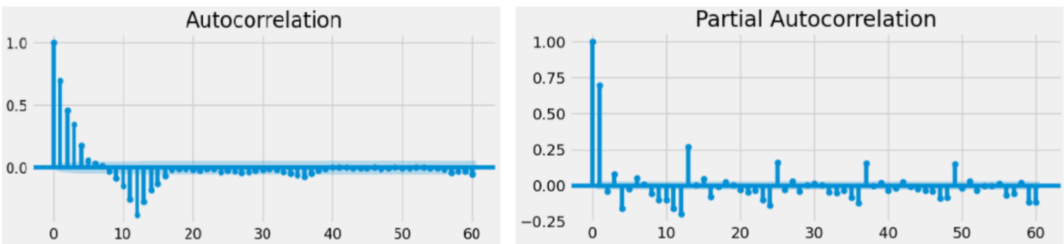

Figure 4.12: ACF and PAF plots for seasonal differenced two-hourly data.

4.6.4 *Neural Network Hyperparameters*

The number of neurons and the layers are essential parameters to choose an appropriate NN architecture in the time series forecasting process which are to be set for the model training and forecasting processes. The number of neurons depends on the number of previously observed values for input layers and forecasted values for output layers. It is decisive to define an appropriate set of hyperparameters in the NN to train and get better forecasting. In general, there is no rule to set the number of neurons in hidden layers and choose the number of hidden layers. It depends on the complexity of the process. Usually, more than one layer is used in deep learning processes [8]. In addition, the number of hidden layers affects the model performance and the burden on the computational time. But, by increasing the number of layers, a better forecasting accuracy may be obtained [19].

In this study, we summarized hyperparameters in Table 4.3 for training and forecasting of the variants of NN model MLP, RNN, and LSTM. We set the number of epochs, batch size and neurons per layer fixed for each NN model. In addition, for model performance comparison, the models are trained by changing the number of hidden layers up to four. We used 80% of the dataset to train and 20% to test or validate the models.

Table 4.3: Neural network hyperparameters.

No.	Parameters	Settings
1	No. of hidden layers	1 to 4
2	No. of epochs	100
3	No. of batch size	73
4	Loss function	MSE
5	Optimization algorithm	Adam
6	Activation function	ReLU
7	No. of neurons per layer	Single hidden layer(512,128,1)
		Two hidden layers(512,128,64,1)
		Three hidden layers(512,128,64,32,1)
		Four hidden layers(1024,512,128,64,32,1)

4.7 MODEL PERFORMANCE EVALUATION

The performance of a model has to be evaluated with an unseen or test dataset to see how the model can predict the output on the test dataset. The prediction error is defined as the difference between the actual and predicted values $e(t) = Z(t) - \hat{Z}(t)$. Several forecasting metrics are used to measure the accuracy of prediction, commonly RMSE, MAE, and MAPE are used to evaluate time series models [8]. The accuracy metrics are mathematically expressed as in equation 4.8, 4.9 and 4.10.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Z(t) - \hat{Z}(t))^2} \tag{4.8}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|Z(t) - \hat{Z}(t)| \tag{4.9}$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|Z(t)-\hat{Z}(t)\right|}{Z(t)} \times 100\% \tag{4.10}$$

There are no best accuracy metrics, except that one considering their interpretation and application. RMSE and MAE are scale-dependent, minimizing will provide forecast of mean and median respectively [18]. The RMSE can be used to identify high-level error values and sensitive to outliers due to the quadratic error value. MAE provides average values within a given range of predicted values or the dispersion of errors. The MAPE is scale-independent, easily interprets percentage errors, and is used to compare more than one series. It has a drawback that the actual value of the data must be greater than zero otherwise it becomes undefined [8] [39] [22] [23]. Moreover MAPE provides higher weights for negative errors and very high values for actual observation closest to zero [18].

In this study, the performance accuracy metrics RMSE, MAE and MAPE are used to evaluate and compare both the statistical and neural network models based on the prediction error. The best model is selected from the performance accuracy measurement considering the least score of RMSE and MAE, besides the MAPE is applied for model comparison.

5 RESULTS AND DISCUSSIONS

This study aims to forecast the VSAT data traffic based on the past time series data observations using multiple time series models. Then the best accurate model is selected to be used as an input for network planning and optimization as per the forecasting performance measured with RMSE, MAE, and MAPE accuracy metrics. In this section, the result of the thesis is presented.

5.1 FORECASTING MODEL RESULTS

In this study, we have selected exponential smoothing, autoregressive and neural network forecasting methods which are frequently used by many researchers for modeling the time series data. We present the result of the selected forecasting models obtained from manual parameter tuning in model training and testing by using 20% of the dataset considering different parameters, the measured accuracy error, and the computational time. In addition, the best models are summarized for comparison based on the result of RMSE, MAE, and MAPE accuracy metrics and two types of dataset applied for analysis.

5.1.1 *Exponential Forecasting Models Result*

The SES model: uses a level smoothing parameter. In this thesis, the level parameter is varied with a step size of 0.11 in the model training process and the models are tested with 20% of the dataset. So that the result of the forecasting models are listed in Table 5.1.

Table 5.1: SES forecasting model results.

α	**RMSE**	**MAE**	**MAPE**
0.00	0.2637	0.2332	20.0454
0.11	0.2117	0.1843	15.7392
0.22	0.1925	0.1668	14.1211
0.33	0.1715	0.1467	12.1889
0.44	0.1618	0.1359	11.0502
0.55	0.1590	0.1315	10.4835
0.66	0.1599	0.1300	10.1849
0.77	0.1628	0.1299	10.0296
0.88	0.1672	0.1307	9.9555
0.99	0.1724	0.1322	9.9567

From the Table 5.1, the minimum forecasting error with RMSE=0.1590, and MAE=0.1315 is obtained. Hence, the SES with $\alpha = 0.55$ is the best from all listed models.

The Holt's Trend forecasting method (DES): with damped additive trend option ϕ=0.8, level α and trend β parameters are chosen for model training by varying the two parameters with a step size of 0.11. The result of the model forecasting is listed in Table 5.2 which are the best combination of α and β parameters.

Table 5.2: DES, Holt's damped additive trend forecasting results.

α	β	**RMSE**	**MAE**	**MAPE**
0.12	0.45	0.1608	0.1346	10.8911
0.12	0.56	0.1629	0.1302	10.0591
0.23	0.12	0.1657	0.1405	11.5523
0.23	0.23	0.1608	0.1300	10.1323
0.34	0.01	0.1681	0.1432	11.8315
0.34	0.12	0.1603	0.1300	10.1580
0.45	0.01	0.1602	0.1339	10.8058
0.56	0.01	0.1591	0.1305	10.3250
0.67	0.01	0.1612	0.1298	10.0928
0.78	0.01	0.1653	0.1303	9.9780

From the Table 5.2, the minimum forecasting error with RMSE=0.1591, and MAE=0.1305 is obtained. So that the DES with α=0.56 and β=0.01 is the best model.

The Holt Winter's forecasting method (TES): we have chosen the model with damped additive trend ϕ=0.8 and multiplicative seasonal options. The forecasting result is shown in Table 5.3 by varying three order parameters level α, trend β and seasonal γ with a step size of 0.11 in model training. The results are the best combination of α, β and γ parameters.

Table 5.3: TES, Holt Winter's forecasting results.

α	β	γ	**RMSE**	**MAE**	**MAPE**
0.12	0.56	0.34	0.1290	0.0860	6.9930
0.12	0.56	0.56	0.1260	0.0854	6.8763
0.12	0.67	0.34	0.1155	0.0872	6.9383
0.23	0.23	0.45	0.1330	0.0897	7.2966
0.23	0.34	0.12	0.1097	0.0762	6.1459
0.34	0.12	0.56	0.1328	0.0913	7.4235
0.34	0.12	0.89	0.1328	0.1052	8.5465
0.56	0.12	0.12	0.1241	0.0925	7.1789
0.67	0.01	0.12	0.1319	0.0940	7.6990
0.89	0.01	0.45	0.1182	0.0901	7.2360

From the Table 5.3, the minimum forecasting error with RMSE=0.1097, and MAE=0.0762 is obtained. So that the TES with α=0.23, β=0.34 and γ=0.12 is the best from other listed models.

5.1.2 *Autoregressive Forecasting Models Result*

The ARIMA model: uses three order parameters p, d and q. From Section 4.6, the ADF and KPSS statistical tests, we confirmed that differencing the series d=1 is needed to make stationary. In addition, using the ACF and PACF we estimated the AR and MA model with optimal parameters p=0,1,2 and q=0,1,2. Therefore, the order parameters p and q are varied in the model training process. Table 5.4 shows the result of the ARIMA model with all combination of p, d and q order parameters.

Table 5.4: ARIMA forecasting model results.

p	d	q	RMSE	MAE	MAPE
0	1	0	0.1729	0.1323	9.9601
0	1	1	0.1899	0.1404	10.3441
0	1	2	0.1986	0.1463	10.7195
1	1	0	0.2095	0.1550	11.3213
1	1	1	0.1941	0.1431	10.5137
1	1	2	0.2021	0.1754	14.9250
2	1	0	0.1769	0.1338	10.0062
2	1	1	0.1693	0.1312	9.9434
2	1	2	0.1671	0.1306	9.9526

From the Table 5.4, the minimum forecasting error with RMSE=0.1671, and MAE=0.1376 is obtained. So that the ARIMA(2,1,2) is the best from other listed models.

The SARIMA model: has seasonal and non-seasonal order parameters including the seasonal period SARIMA(p,d,q)x(P,D,Q)x[s]. From Section 4.6, having the two-hourly resampled data, we identified the need of seasonal stationary D=1 by using ADF and KPSS tests. In addition, the order parameters are estimated by using the ACF and PACF tests and the optimal parameters are p=0,1,2, q=0,1,2, d=1, D=1, P=0,1,2, Q=0,1 and s=12. These order parameters are varied in the model training process. Table 5.5 shows the result of the SARIMA model with the best combination of p, d, q, P, D, Q and s order parameters.

Table 5.5: SARIMA forecasting model results.

p	d	q	P	D	Q	s	RMSE	MAE	MAPE
0	1	0	0	1	1	12	0.1211	0.0894	7.3645
0	1	1	0	1	1	12	0.1266	0.0981	8.1103
0	1	2	0	1	1	12	0.1221	0.0877	7.2016
1	1	0	0	1	1	12	0.1239	0.0947	7.8246
1	1	1	0	1	1	12	0.1321	0.0900	7.3335
1	1	2	0	1	1	12	0.1264	0.0873	7.1267
2	1	0	0	1	1	12	0.1251	0.0960	7.9340
2	1	1	0	1	1	12	0.1250	0.0961	7.9422
2	1	1	2	1	0	12	0.1401	0.1037	8.5245
2	1	2	0	1	1	12	0.1276	0.0877	7.1546

From the Table 5.5, the minimum forecasting error with RMSE=0.1211, and MAE=0.0894 is obtained. So that the SARIMA(0,1,0)x(0,1,1)x[12] is the best from other listed models.

5.1.3 *Neural Network Forecasting Models Result*

Following the hyperparameter setting in Section 4.6 and Table 4.3, we have made model training by changing the number of hidden layers up to four as a criteria to get best combination of parameters for each NN model. Therefore, the result of ANN forecasting models are listed in Table 5.6.

Table 5.6: ANN forecasting model results.

ANN(I,H,O) Models	RMSE	MAE	MAPE
MLP(512,128,1)	0.0611	0.0442	3.5243
MLP(512,128,64,1)	0.0634	0.0465	3.6482
MLP(512,128,64,32,1)	0.0720	0.0549	4.4596
MLP(1024,512,128,64,32,1)	0.0617	0.0450	3.5115
RNN(512,128,1)	0.0638	0.0474	3.7159
RNN(512,128,64,1)	0.0603	0.0433	3.3838
RNN(512,128,64,32,1)	0.0639	0.0465	3.6019
RNN(1024,512,128,64,32,1)	0.0733	0.0554	4.2809
LSTM(512,128,1)	0.0634	0.0468	3.7400
LSTM(512,128,64,1)	0.0704	0.0531	4.2725
LSTM(512,128,64,32,1)	0.0643	0.0474	3.7348
LSTM(1024,512,128,64,32,1)	0.0660	0.0493	3.9474

According to Table 5.6, from the MLP models, the minimum forecasting error RMSE=0.0611 and MAE=0.0442 is obtained when one hidden layer MLP(512,128,1) is applied. In addition, from the RNN models, the minimum error RMSE=0.0603 and MAE=0.0433 is obtained in case of two hidden layers RNN(512,128,64,1) are set. Finally, from LSTM models, the minimum error RMSE=0.0634 and MAE=0.0468 is obtained when single hidden layer LSTM(512,128,1) is used.

Furthermore, Figure 5.1, shows the different performance result of the ANN models with respect to the number of hidden layers as per the measurement of MAE forecasting metrics.

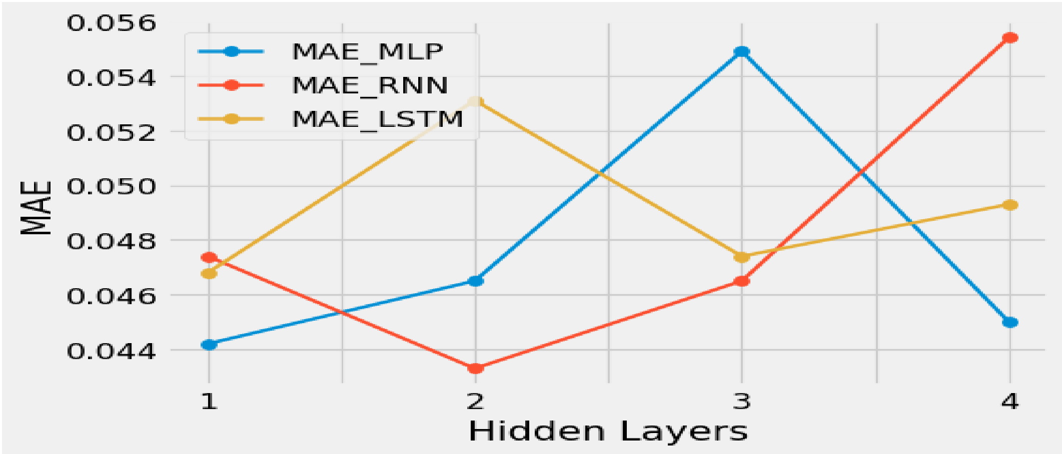

Figure 5.1: The ANN models performance for hidden layers.

From Figure 5.1, ANN models produced an increasing forecasting error when the number of hidden layers were increasing, but the errors were declining for the MLP model. High forecasting errors observed in RNN model when four hidden layers were set. In addition, the least errors scored when two hidden layers were used in RNN model, and when single hidden layer was used in MLP model. As a result, the RNN model outperformed other models and produced the least errors when two hidden layers applied.

5.1.4 *Predictability Test Result*

Time series data have characteristics of short and long-term dependency (self-similarly behavior). The Hurst exponent index H is the common statistical method to identify the trend persistence and evaluate the predictability. Often the Hurst exponent H ranges between (0, 1), tells the characteristics of the time-series data, is measured by the rescaled range (R/S) methods [33] [34] [35] [11] [7].

- If $0 < H < 0.5$: - mean-reverting (anti-trend). As H near to 0 indicates anti-persistence (short memory), strong non-linearity and weak predictability.

- If H = 0.5: - random with no similarity and unpredictable.
- If 0.5 < H < 1: - self-similar with long-range dependence (trendiness). As H is near to 1 indicates persistence (long memory) and predictable.
- If H > 1: - non-stationary (periodic) patterns and ensures more predictable.

The slop of a line plotted between the (R/S)s and S on the log-log scale will determine the Hurst exponent. The R/S method is denoted by equation 5.1.

$$(R/S)_n = Cn^H \tag{5.1}$$

Where n is the length of the series, R is rescaled range, S is the standard deviation, H is Hurst exponent and C is a constant.

In this thesis, we used Hurst exponent to test the data traffic dataset to have an insight on the behaviour of the dataset and for further analysis as shown in Figure 5.2 resulting H = 0.1013 which indicates anti-persistence or week-trend behavior, as well as having a short memory and less predictable.

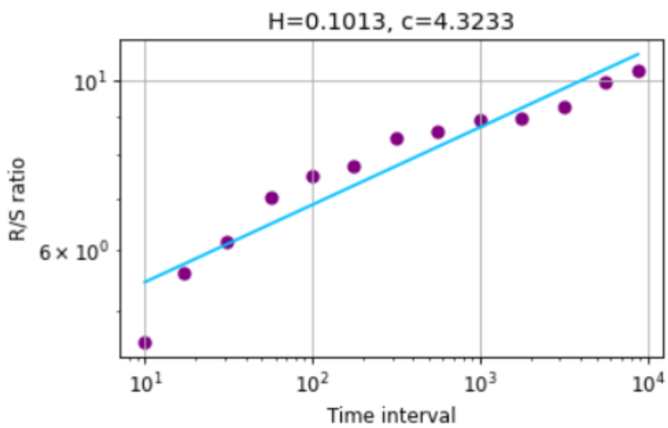

Figure 5.2: Hurst exponent index H with R/S method.

5.2 COMPARISON OF FORECASTING MODELS

In this section, we have summarized all the selected models having the minimum forecasting errors from the result of model testing including the computational time taken during model training. The summary of the best models is shown in Table 5.7.

Table 5.7: Summary of best forecasting models.

Models[DS-A]	**RMSE**	**MAE**	**MAPE**	**CompTime[min]**
SES(α=0.55)	0.1590	0.1315	10.4835	0.0039
DES(α=0.56,β=0.01)	0.1591	0.1305	10.3250	0.1842
TES(α=0.23,β=0.34,γ=0.12)	0.1097	0.0762	6.1459	3.0005
ARIMA(2,1,2)	0.1671	0.1306	9.9526	0.1623
SARIMA(0,1,0)x(0,1,1)[12]	0.1211	0.0894	7.3645	5.8996
MLP(512,128,1)	0.0611	0.0442	3.5243	0.2593
RNN(512,128,64,1)	0.0603	0.0433	3.3838	0.6676
LSTM(512,128,1)	0.0634	0.0468	3.7400	4.5115

From Table 5.7 referring to the RMSE metric, we have found that the statistical time series models produced the highest errors. The Holt Winter's (TES) and SARIMA models have the minimum errors with high computational time and performed better due to the seasonality behavior of the dataset. Nevertheless, TES is better than the SARIMA model and the SARIMA model is computationally more expensive. The other models SES, DES and ARIMA require less computational time, however their performance is poor as the data has anti-trend behavior. Regarding to the ANN time series models produced the lower errors and outperformed the statistical models. The RNN model performed the best as the dataset has a short-range dependency and best suited for such type of series. The LSTM model is the last and requires higher computational time. LSTM is more suitable for series having a long-range dependency. In general, the machine learning models take less computational time than the statistical models.

Moreover, from the Table 5.7 we compared the time series models as per their MAPE scores as shown in Figure 5.3. From statistical models, the TES improves the forecasting accuracy by 41.38% than SES and 16.55% than SARIMA models. Similarly, form the ANN models, the the RNN model improves the forecasting performance by 3.99% than MLP and 9.52% than LSTM. As a result, the forecasting performance of the RNN model has an improvement of 44.94% than the statistical TES model. Finally, the RNN model with two hidden layers RNN(512,128,64,1) is found the best and more accurate than all selected best forecasting models. Therefore, it is recommended for predicting the VSAT network traffic in the future.

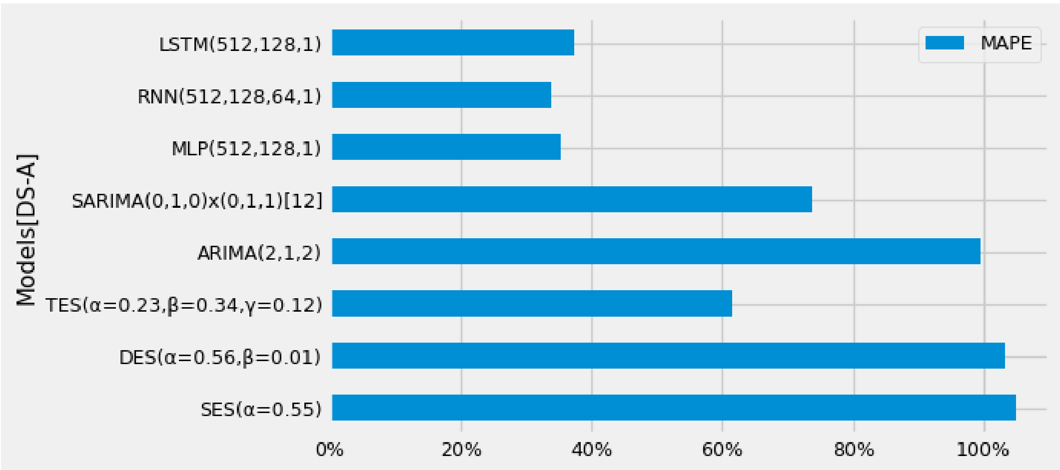

Figure 5.3: Performance accuracy of forecasting models.

Furthermore, we have tested the models with pre-processed two datasets to see their effects on the performance of the models. The first data labeled as 'DS-A2' is without filtering the outliers and the second data labeled as 'DS-B' is with the weekly outage time without considering them as missing values and 'DS-A' is with the outage time considered as a missing values. The models are compared based on the forecasting accuracy of MAPE metrics for each dataset shown in Figure 5.4.

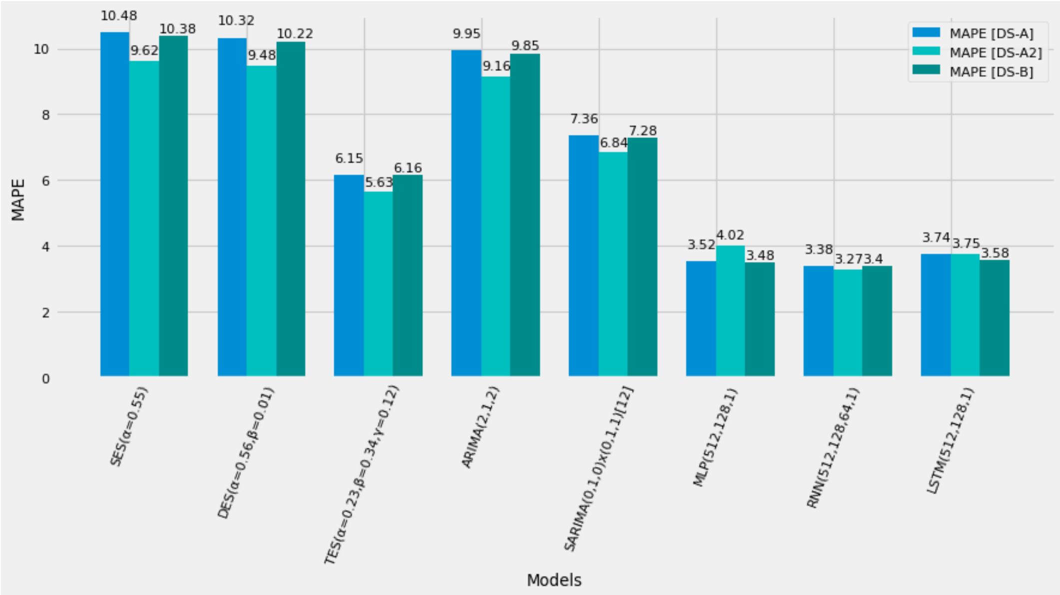

Figure 5.4: Models performance on different datasets.

In this thesis, based on the results shown in Figure 5.4, there is a slight difference in the model performances for both statistical and machine learning between the datasets used with and without estimating the weekly outage times as a missing values by LR interpolation. In addition, the ANN models exhibited more errors for the dataset without filtering the outliers than the statistical models, therefore, they are more sensitive to few outliers in the dataset than statistical models.

6

CONCLUSION AND RECOMMENDATION

6.1 CONCLUSION

This study aimed to build models for the time series VSAT network data traffic by evaluating the forecasting performance of the statistical and machine learning models. Then the best model that can forecast the VSAT data traffic with the minimum error was to be identified by comparing the model performances.

In this thesis, the VSAT data traffic was collected from the edge router (ASR) at the station, and the data preprocessing was done to make the dataset consistent as well as being used as an input for model training and testing. In the data analysis process, the time series data was passed through pattern visualization, component decomposition, and predictability test to have an insight into the behavior of the series and the parameters to take into consideration. In addition, statistical tests were performed to estimate the order parameters of ARIMA and SARIMA models by using ACF and PACF plots, besides ADF and KPSS tests were also used to decide the need for differencing for the series to make it stationary. Moreover, the smoothing parameters were set within the specified range of (0,1) for the exponential smoothing methods SES, DES and TES. Regarding to the ANN models, the hyperparameters were set for model training. Then the best combination of order parameters for exponential smoothing, autoregressive, and ANN models were tuned by the process of model training and tested based on forecasting error measurements. The performance of the models was measured by the RMSE, MAE, and MAPE accuracy metrics. From the summary of results, the best-performing models were selected for model analysis and comparison.

From the result of forecasting performance process and the list of best performing models, the following conclusions can be drawn.

1. From the statistical models SES, DES, and ARIMA exhibited more forecasting errors since the time-series data showed anti-trend or non-persistence characteristics. The other TES and SARIMA models produced fewer errors and performed better as the series had seasonality behavior, but they required high computational time.

2. The ANN modes produced different performances with regard to the number of hidden layers. For the case of increasing the number of hidden layers, the ANN models generated more errors, but MLP produced a declining errors. From all ANN models, the RNN exhibited the least errors as the time series had a short-range dependency.

3. For the time series data, the performance of ANN models were found better than the statistical models. In terms of computational time, the ANN models took very less time than statistical models.

4. From the summary of the forecasting result with MAPE metrics, the RNN model was found the best model and improved the forecasting performance by 44.94% than the statistical TES model.

6.2 RECOMMENDATION

In this thesis, the overall result showed that the RNN model with two hidden layers is found the best forecasting model. Therefore, the RNN model is recommended to Ethio Telecom to forecast the VSAT data traffic demand

and can be used as input for the network planning and optimization tasks. Moreover, it will enhance resource utilization, QoS and increase customer satisfaction.

Furthermore, this thesis can be considered as a baseline or benchmark, and further expanded by future works as outlined below.

- This study was conducted on a yearly network traffic data from a single VSAT network. Therefore, including the other aggregated VSAT network data will increase the data size. So that deep learning neural network models can be used for the big data size to achieve better performance.
- Combining the statistical and neural network models can be used to get a better performance in the model comparison process, including the auto optimization model training methods.
- In this study, the forecasting methods were performed for univariate time series data. As a result, by taking into account additional variables, multivariate models can be used to forecast VSAT data traffic.

BIBLIOGRAPHY

[1] Muhammad Faisal Iqbal, Muhammad Zahid, Durdana Habib, and Lizy Kurian John. "Efficient prediction of network traffic for real-time applications." In: *Journal of Computer Networks and Communications* 2019 (2019). DOI: https://doi.org/10.1155/2019/4067135.

[2] Ethio Telecom. *Fixed Broadband Services*. [Online, accessed on 04-Jun-2021]. 2021. URL: https://www.ethiotelecom.et/.

[3] Ethio Telecom. *Broadband VSAT Customer Data*. Infrastructure transport network O&M, Satellite hub and VSAT section, Addis Ababa. 2021. URL: https://www.ethiotelecom.et/.

[4] Murali Krishna P, Vikram M Gadre, and Uday B Desai. *Multifractal Based Network Traffic Modeling*. 1st Ed. New York: Springer Science+Business Media, LLC, 2003. ISBN: 978-1-4615-0499-3. DOI: 10.1007/978-1-4615-0499-3.

[5] Sheetal Thakare, Anshuman Pund, and MA Pund. "Network Traffic Analysis, Importance, Techniques: A Review." In: *2018 3rd International Conference on Communication and Electronics Systems (ICCES)*. IEEE. 2018, pp. 376–381. DOI: 10.1109/CESYS.2018.8723955.

[6] Thomas M Chen. "Network traffic modeling." In: *The handbook of computer networks*. Vol. 3. Wiley Hoboken, NJ, 2007, p. 156. ISBN: 978-0-471-78460-9.

[7] Samuel Medhn, Bethelhem Seifu, Amel Salem, and Dereje Hailemariam. "Mobile data traffic forecasting in UMTS networks based on SARIMA model: The case of Addis Ababa, Ethiopia." In: *2017 IEEE AFRICON*. IEEE. 2017, pp. 285–290. DOI: 10.1109/AFRCON.2017.8095496.

[8] Oleg Ostashchuk. "Time series data prediction and analysis." MSc Thesis. Czech Technical University in Prague Faculty of Electrical Engineering Department of Computer Science, 2017.

[9] Atif Mahmood, Miss Laiha Mat Kiah, Muhammad Reza Z'aba, Adnan N Qureshi, Muhammad Shahreeza Safiruz Kassim, Zati Hakim Azizul Hasan, Jagadeesh Kakarla, Iraj Sadegh Amiri, and Saaidal Razalli Azzuhri. "Capacity and frequency optimization of wireless backhaul network using traffic forecasting." In: *IEEE Access* 8 (2020), pp. 23264–23276. DOI: 10.1007/s12652-020-02209-2.

[10] Getinet Tesfaye. "Hybrid SARIMA-ELM-based Data Traffic Forecasting: The Case of UMTS Network in Addis Ababa, Ethiopia." MSc Thesis. Addis Ababa Institute of Technology (AAiT), School of Electrical and Computer Engineering, 2018.

[11] Zhongda Tian. "Network traffic prediction method based on wavelet transform and multiple models fusion." In: *International Journal of Communication Systems* 33.11 (2020), e4415. DOI: 10.1002/dac.4415.

[12] Gerard Maral. *VSAT networks*. 2nd Ed. Southern Gate, UK: John Wiley & Sons Ltd, 2004. ISBN: 0-470-86684-5.

[13] Gerard Maral, Michel Bousquet, and Zhili Sun. *Satellite communications systems: systems, techniques and technology*. 6th Ed. 111 River Street, Hoboken, NJ 07030, USA: John Wiley & Sons Ltd, 2020.

[14] Khalid Abdalrazig Ibrahim Hassan and Amin Babiker A/Nabi Mustafa. "VSAT Network Overview." In: *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)* Volume 10, (2015), pp. 18–24. ISSN: 2278-2834. DOI: 10.9790/2834-10131824.

[15] Anil K Maini and Varsha Agrawal. *Satellite technology: principles and applications*. 2nd Ed. Southern Gate, UK: John Wiley & Sons Ltd, 2011. ISBN: 978-0-470-66024-9.

[16] HNS. *HN System System Overview Release 5.8*. Revision A. [System Manual]. Hughes Network Systems LLC. 11717 Exploration Lane Germantown Maryland 20876, 2008. URL: https://www.hughesnet.com/.

[17] Nuri Ucar. "Comparison of the forecast performances of linear time series and artificial neural network models within the context of Turkish inflation." MSc Thesis. Bilkent University, 2001.

[18] Emma Wallentinsson. "Multiple Time Series Forecasting of Cellular Network Traffic." MSc Thesis. Linkoping University, Department of Computer and Information Science, 2019.

[19] Begameder Tamene. "Video Streaming Data Traffic Prediction by Using LSTM Model." MSc Thesis. Addis Ababa Institute of Technology (AAiT), School of Electrical and Computer Engineering, 2020.

[20] George E P Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time Series Analysis: Forecasting and Control*. 5th Ed. Hoboken, New Jersey.: John Wiley & Sons, Inc., 2016. ISBN: 978-1-118-67502-1.

[21] Peter J Brockwell and Richard A Davis. *Introduction to time series and forecasting*. 3rd Ed. Springer International Publishing, Switzerland, 2016. ISBN: 978-3-319-29854-2. DOI: 10.1007/978-3-319-29854-2.

[22] Chris Chatfield. *Time-series forecasting*. Department of Mathematical Sciences, University of Bath, UK: Chapman & Hall/CRC press, 2000. ISBN: 1-58488-063-5.

[23] Peter Wallstrom. "Evaluation of forecasting techniques and forecast errors: with focus on intermittent demand." MSc Thesis. Lulea University of Technology, 2009.

[24] Neil Y Yen, Jia-Wei Chang, Jia-Yi Liao, and You-Ming Yong. "Analysis of interpolation algorithms for the missing values in IoT time series: a case of air quality in Taiwan." In: *The Journal of Supercomputing* 76.8 (2020), pp. 6475–6500. DOI: 10.1007/s11227-019-02991-7.

[25] Nikhil Ketkar. *Deep Learning with Python: A Hands-on Introduction*. Bangalore, Karnataka, India: Apress, 2017. ISBN: 978-1-4842-2766-4. DOI: 10.1007/978-1-4842-2766-4.

[26] Wucherl Yoo and Alex Sim. "Time-Series Forecast Modeling on High-Bandwidth Network Measurements." In: *Journal of Grid Computing* 14.3 (2016), pp. 463–476. DOI: 10.1007/s10723-016-9368-9.

[27] Colin Gorrie. *Three ways to detect outliers*. [Online, accessed on 31-Dec-2021]. 2016. URL: http://colingorrie.github.io/tag/outliers.html.

[28] Michael Galarnyk. *Understanding boxplots; towards data science*. [Online, accessed on 31-Dec-2021]. 2018. URL: https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51.

[29] Jason Brownlee. *How to remove outliers for machine learning; machine learning mastery*. [Online, accessed on 01-Jan-2022]. 2018. URL: https://machinelearningmastery.com/how-to-use-statistics-to-identify-outliers-in-data/.

[30] Jason Brownlee. *Better deep learning: train faster, reduce overfitting, and make better predictions*. Ed. v1.3. eBook. Machine Learning Mastery, 2018.

[31] Chris. *How to Normalize or Standardize a Dataset in Python?* [Online, accessed on 19-Jul-2021]. 2020. URL: https://www.machinecurve.com/index.php/2020/11/19/how-to-normalize-or-standardize-a-dataset-in-python/.

[32] Keita Miyaki. *Time Series Split with Scikit-learn*. [Online, accessed on 10-Jul-2021]. 2019. URL: https://medium.com/keita-starts-data-science/time-series-split-with-scikit-learn-74f5be38489e.

[33] Rajandran R. *Hurst Exponent – Checking for Trend Persistance*. [Online, accessed on 19-Aug-2021]. 2020. URL: https://www.marketcalls.

in/python/hurst-exponent-checking-for-trend-persistance-python-notebook.html.

[34] Christos Katris and Sophia Daskalaki. “Comparing forecasting approaches for Internet traffic.” In: *Expert Systems with Applications* 42.21 (2015), pp. 8172–8183. DOI: 10.1016/j.eswa.2015.06.029.

[35] Sivapragasam Chandrasekaran, Saravanan Poomalai, Balamurali Saminathan, Sumila Suthanthiravel, Keerthi Sundaram, and Farjana Farveen Abdul Hakkim. “An investigation on the relationship between the Hurst exponent and the predictability of a rainfall time series.” In: *Meteorological Applications* 26.3 (2019), pp. 511–519. DOI: 10.1002/met.1784.

[36] Kefentse Mokgolodi. “Traffic prediction in cloud computing using time series models.” MSc Thesis. University of Botswana, Department of Computer Science, 2016.

[37] Josef Perktold, Skipper Seabold, and Jonathan Taylor. *Stationarity and detrending (ADF/KPSS)*. [Online, accessed on 03-Aug-2021]. 2019. URL: https://www.statsmodels.org/dev/examples/notebooks/generated/stationarity_detrending_adf_kpss.html.

[38] Amel Salem. “SARIMA Based Data Traffic Forecasting; The Case of UMTS Network in Addis Ababa.” MSc Thesis. Addis Ababa Institute of Technology (AAiT), School of Electrical and Computer Engineering, 2016.

[39] Tianfeng Chai and Roland R Draxler. “Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature.” In: *Geoscientific model development* 7.3 (2014), pp. 1247–1250. DOI: 10.5194/gmd-7-1247-2014.

A

APPENDIX

A.1 MANUSCRIPT FOR PUBLICATION