Addis Ababa University

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

Telecommunication Engineering Graduate Program

Time Series Forecasting for Incoming Call Volume using LSTM: the case of Ethio Telecom Call Center

By: Mebrhit Hailay

Advisor: Dr. -Ing. Dereje Hailemariam

A Thesis Submitted to the School of Graduate Studies of Addis Ababa University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Telecommunication Engineering

January, 2022

Addis Ababa, Ethiopia

Addis Ababa University

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

Telecommunication Engineering Graduate Program

MSc Thesis on:

Time Series Forecasting for Incoming Call Volume using LSTM: the case of Ethio Telecom Call Center

By: Mebrhit Hailay Abraha

Approval by Board of Examiners

______________________________ _____________

Chair/ School Dean Signature

Dr.-Ing. Dereje Hailemariam ______________
Advisor Signature

Ephrem Teshale (PhD) ______________
Examiner 1 Signature

Surafel Lemma (PhD) ______________
Examiner 2 Signature

Declaration

I, the undersigned, declare that this thesis is my original work, has not been presented for a degree in this or any other university, and all sources of materials used for the thesis have been fully acknowledged.

Mebrhit Hailay ______________

Name: Signature:

This thesis has been submitted for examination with my approval as a university advisor.

Dr.-Ing. Dereje Hailemariam ______________

Advisor's Name Signature

Abstract

An accurate forecasting of incoming call volume traffic is key for the operational planning of an inbound call center service. By forecasting, the correct number of incoming calls volumes traffic, can determine staffing and scheduling levels, improve service requirements, and meet customer satisfaction. Currently, in Ethio Telecom inbound type of call center service, the averaging method is used for forecasting incoming call volume traffic. But there is a problem with the averaging method of forecasting the number of incoming calls volume traffic in a call center that cannot handle the trend and seasonality fluctuation. The purpose of this study is to build a model that forecasts incoming call center call traffic using Ethio Telecom call center historical data.

In this thesis, we proposed time series forecasting model for forecasting the incoming call volume traffic. We have used two univariate time series techniques, namely SARIMA and LSTM. Nine months of incoming call center call traffic data is collected from Ethio Telecom. Finally, experimental result indicates that the LSTM model has 24.6% of RMSE improvement of forecasting error compared to the SARIMA model. The overall results of this research work demonstrate that the LSTM model is an effective method for predicting incoming call volume traffic to reflect temporal patterns. Such accuracy is vital to provide a better call center resource allocation for optimization staffing and scheduling problem.

Keywords – *call center, time series, forecasting, Staffing, scheduling, statistical model, deep neural network, LSTM*

Acknowledgments

I would like to thank my advisor Dr. - Ing. Dereje Hailemariam for his genuine pieces of advice throughout the thesis work and his suggestions are valuable and helped me shape my work.

I would like to thank Ethio telecom staff members, especially Ato. Tselotehaimanot Molla, Mr. Tefetawit Gregzabeher and Mr. Haimanot Hilezgy as they were very cooperative in providing the necessary data for this work and as they were very encouraging throughout the whole process.

Last but not least I also want to give special thanks to my husband, Eng. Addisalem Tafre and my beloved baby boy Adonay Addisalem and to all my parents too.

Finally, I would like to thank my friends, Gebremeskel Aregay, Desalegn Medhin, G/medhin W/Mariam and Tewelde Mezgebo for giving me support, strength, and continuous encouragement throughout my work and through the process of researching and writing the thesis. This achievement would not have been possible without you. Thanks to all my family and my friends.

Table of Contents

List of Figures

List of Tables

Acronyms

ACF	Auto correlation Function
ADF	Augmented Dickey Fuller
AIC	Akaike information criterion
ANN	Artificial Neural Networks
AR	Autoregressive model
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
CRM	Customer Relationship Management
CTI	Computer Telephony Integration
DL	Deep learning
IPCC	Internet Protocol Contact Center
IVR	Interactive Voice Response
LSTM	Long-Short Term Memory
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
PACF	Partial Autocorrelation Function
QoS	Quality of Service
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SARIMA	Seasonal Auto Regressive Integrated Moving Average
SM	Statistical Model
UAP	Universal Access Platform
VIP	Very Important Person
VoIP	Voice over Internet Protocol

CHAPTER ONE

Introduction

Many businesses are establishing their own call center service departments these days. With technological advancements, managing call center service efficiently and effectively is critical in terms of economic interest, providing quality of service (QoS) and improving customer experience [1]. Call center is a service providing department which is used as a frontline that allows customers to request and get support directly from the company. It is one of the most important forms of communication between businesses and their customers. Call centers can be used in different business areas such as banks, airlines, healthcare, and telecommunications to handle a high volume of customer call requests [2] [3].

The increasing number of businesses encourages companies to have their own call center to provide the best possible services and information to their customers [4][5]. Customers save significant time by using call centers efficiently, while businesses run significantly more efficient operations. In order to improve customer satisfaction, call center managers face numerous challenges in providing high-quality service while keeping operating costs low. Because managers are continually presented with tremendous uncertainty, effective call center management is a difficult undertaking. Because it is often time-varying, stochastic, reliant on time periods and call kinds, and frequently influenced by external events, the number of incoming calls is a substantial source of uncertainty. Accurate modeling and forecasting of future call arrival volumes is a complicated subject that is vital for important call center operational decisions [6].

Mainly there are two types of call centers: inbound and outbound call centers. Inbound call centers are the most common type of call centers, where customers initiate calls and call center agents respond to those calls. Outbound call centers are another type of call center in which the company makes outgoing calls to customers [3].

The data generated from call centers have a significant amount that should be analyzed and used for decision-making purposes in order to determine appropriate staff levels, shift schedules, and real-time routing design future demands [7] [8].

To effectively run a call center, call center managers must match call center resources to workload. The first and most important step in forecasting workload effectively is to provide an accurate forecast of future call volumes. A call center manager frequently requires two types of projections for staffing and scheduling [9] [10].

1. Forecast call volumes several days or weeks in advance;
2. Dynamically update the forecast on a given day based on newly available data as new calls arrive throughout the day.

1.1 Background of Ethio Telecom

Ethio Telecom is a telecommunication service provider which operates in Ethiopia. It is the sole telecom operator in the country and is owned by the Government of Ethiopia. Ethio Telecom possesses a huge customer base, and its total number of subscribers has reached 56.2 million in July 2021 [11].

Currently, Ethio Telecom provides voice, internet, data and value added services (VAS) throughout Ethiopia. Ethio Telecom now has several customer service systems in place, including inbound and outbound call center services, as well as multichannel service (Short Message Service, Telegram chat, Facebook chat and Twitter chat). Inbound call center agents handle incoming calls to an Ethio Telecom. Outbound call center agents make outgoing phone calls to customers.

Inbound call center service provides services via free online telephone (994), which is supported by interactive voice response (IVR), and it serves customers in five languages: English, Tigrigna Amharic, Oromiffa, and Somali. Inbound call center services assist customers with questions regarding Ethio Telecom products and services, subscription requirements, billing and associated inquiries, activation of value added services, registration faults, status updates, and follow up. Whereas, outbound call center service is reserved for the company's selected enterprise, very important person(VIP), and key account customers, who are served in both English and Amharic languages.

In Ethio Telecom inbound call center services, incoming calls from customers requesting information are handled using IVR and agent handling methods. Currently, in the agent handled method, the averaging (handwriting) method is used for forecasting future incoming call volumes traffic to be used in determining staff level and shift scheduling of agents. Although it is important to improve service requirements and customer satisfaction, the agent handled method experiences call drop problems i.e., calls are interrupted before they get service. For instance, as shown in figure 1.1 of one day total incoming calls report, 69.1% of the calls get lines and served. Incoming calls are handled based on the needs of the customer, who receives a response via IVR or agent methods.

From the IVR method, 61.39% of customer calls are successfully answered by IVR, while the remaining 38.61% calls are blocked (dropped calls). From the agent method, 41.81% customer calls are successfully responded to by agents, while the remaining 58.19 % calls are blocked.

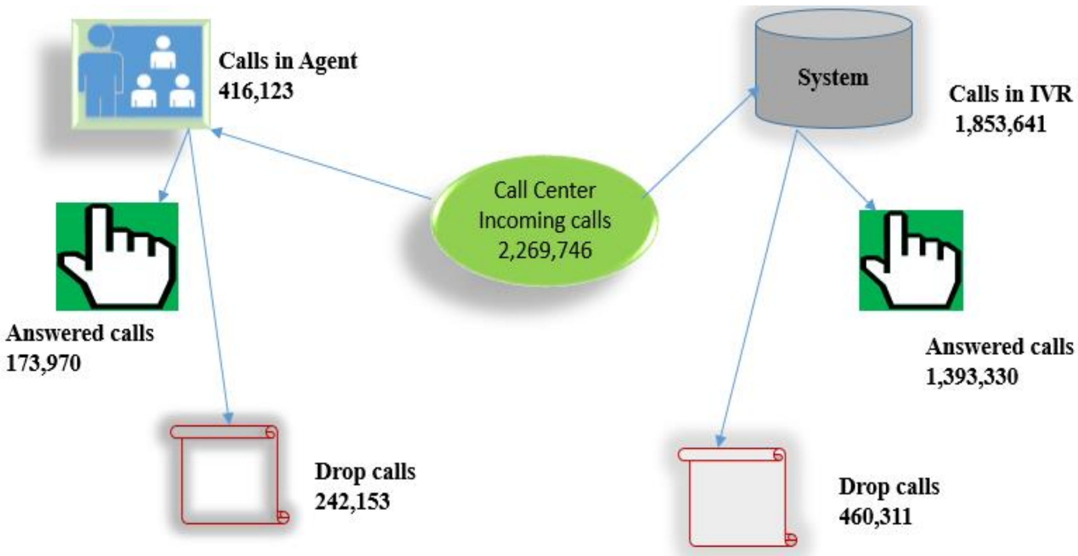


Figure 1.1 sample report of incoming call volume for one day

Generally, in the agent handled method, the number of dropped calls is higher than the number of answered calls. Where as in the IVR method, call success rate is higher than call drop rate. So, this indicates that the calls that are handled by agents have low performance leading to customer dissatisfaction, and need analysis in order to minimize the dropped calls. Besides, Ethio Telecom is currently the only operator in Ethiopia, and the Government has a policy to privatize the telecom sector and has allowed other competitors to enter the market. As a result, customers will have the opportunity to choose operators having good quality of service (QoS). So, to maintain its customers, Ethio telecom must manage call center service operations effectively and efficiently. This helps minimize dropped calls and waiting time leading to improving service requirements and meet customer satisfaction. Therefore, a mechanism that accurately forecasts future incoming call volume traffic is critically required to handle service operations of a call center efficiently.

1.2 Statement of the Problem

In companies such as Ethio Telecom, inbound customer calls to request information are served through call centers, and most of the operational expenditures of call centers are related to workforce. So, efficiently managing labor forces in call center is a critical function. However, managers face a problem of knowing accurate incoming workloads. As a result, managers are not able to properly and efficiently perform staff scheduling i.e., assigning the right personnel at the right time to handle incoming calls. This leads to high operational costs and customer dissatisfaction. To solve these problems, proactive forecast of accurate incoming call volume is required.

Some research works such as [12] and [13] have been conducted to develop a model that forecasts incoming calls of a call center. These researches are based on statistical analysis models that better handle time series data having linear relationships. However, incoming calls of call centers have complex and non-linear behaviors, and their accurate forecast is very critical to efficiently perform staff scheduling. So, an algorithm that better understands the complex and non-linearity relationships of inbound customer calls is required to improve forecast accuracy.

This research work is to develop a model that forecasts incoming call volumes using deep neural network, Long Short Term Memory (LSTM) model to improve call volume forecast accuracy for Ethio Telecom call center.

1.3 Objectives

1.3.1 General Objective

The general objective of this study is to develop time series based forecasting model for incoming call volumes traffic using LSTM algorithm for Ethio Telecom call center.

1.3.2 Specific Objectives

To accomplish the general objective of the study the following specific objectives were targeted.

- Understand the basic call center system design principle.
- Review appropriate forecast models for call traffic volumes.
- Collect, prepare appropriate data and visualize call traffic volumes data from call center service operation of Ethio Telecom.

- Select the best forecasting algorithms that fit into call traffic volumes.
- Implement the selected forecasting algorithms using a python tool.
- Evaluate the performance using forecast error measures like Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE).
- Propose the one with minimum prediction error.

1.4 Scope and Limitation of the Study

The scope of this study is to forecast the number of incoming call volume traffic for inbound call centers using the LSTM model. In addition, compare the results with the statistical forecasting model, SARIMA, and select the best one with the minimum forecast error. In addition, this study addresses the importance of forecasting in the context of call center operational resource planning (staffing and scheduling problem).

1.5 Contribution of the Study

This thesis work can help call center managers to manage resources effectively by minimizing call drop in order to meet the target of the service level agreement of the company. It can also enhance the customer satisfaction by providing service on-time and the revenue of the company can be increased when customers are timely served and services are maintained immediately. The contribution can also be extended as an input for other researchers to conduct a research in different perspectives.

1.6 Methodology

To accomplish the general and specific objectives of this study, we follow the following methods:

- Reviewing various literature, conference papers, journals, white papers, and books related to call traffic volumes forecasting can help understand forecasting methods and the area, as well as gain knowledge on how others have perceived and dealt with the problem.
- Data was collected for 273 days from the inbound call center service IPCC system of Ethio Telecom.

- The collected data was pre-processed, explored, and visualized and then data was divided into two groups for training and testing.
- Implement the SARIMA and LSTM forecasting algorithm using python tool and evaluate the performance using forecast error measure (RMSE, MAE and MAPE)
- Finally, the findings and results are discussed, and recommendations are made.

1.7 Related Work

As call center call traffic volume forecasting is critical, several authors have been searching better call center traffic forecasting methods.

In [6], different statistical time series methods for forecasting future call volumes in call centers are evaluated and compared. The main motivation of the study is to solve the challenges faced by call center manager such as the problem of determining appropriate staffing levels, the problem of scheduling (and re-scheduling) the available pool of agents based on updated forecasts and routing incoming calls in real-time to available agents in the major company of Canada. In the study, they consider different forecasting models including the Fixed-Effects (FE) Model, the Gaussian Linear Mixed Effects (ME) Model, the Holt-Winters (HW) Smoothing model and the "Top-Down" (TD) Approach. Then they extend the mixed model into a bivariate mixed model which exploits the correlation between the arrival processes of two separate call types and compare their forecasting accuracy. Finally, they show result that the bivariate mixed model yields more accurate forecasts than mixed model.

In the study [12], the authors have forecasted the number of incoming call center calls. To study the dynamic time series three forecast methods namely, Moving Averages method, Simple Exponential Smoothing method and Additive Holt-Winters method have been used. The authors also evaluated the forecasting accuracy of selected models with real-life call center data. The Additive Holt-Winters method is best performed than the others two methods used.

The study in [14] mainly evaluates univariate time series methods: seasonal autoregressive integrated moving average (SARIMA) modeling; robust exponential smoothing based on exponentially weighted least absolute deviations regression; dynamic harmonic regression, which is a form of unobserved component state-space modeling for forecasting intraday arrivals, periodic autoregressive modeling; and an extension of Holt-Winters exponential smoothing for the case of two seasonal cycles, for lead times from half-hour ahead to two weeks ahead. And the main focus of this paper is on the comparison of univariate time series forecasting methods for recorded arrival data from five call center of a bank in United Kingdom.

The authors in [15] have proposed a Big Data framework for forecasting incoming calls to the mobile call center. They conducted data analysis using historical data of incoming calls from 2015-2018 years and defined two models and flow diagrams. Then they develop web applications for ease to use. Finally, they conclude the proposed framework is to provide an accurate forecasting method, easy to use and interact with the forecast application service and save time to forecast than the existing methods.

The authors in [16] investigated time series forecasting methods to effectively manage the bank's workforce call center, and the seasonal moving average (SMA) method is used for systematic evaluations, as are artificial neural networks (ANNs) for forecasting intraday call arrivals. To improve short-term forecasting accuracy, a hybrid method is proposed that combines the strengths of the seasonal moving average method and nonlinear data-driven artificial neural networks (ANNs).

The study in [17] determines the optimal number of agents in contact center service based on downtime and administrative task duration parameters. The authors have proposed a mathematical model for contact centers and a six step estimation model to compute the optimal number of contact. Finally, they concluded the Erlang C formula tool is suitable for modeling the QoS parameters of contact centers.

In conclusion, all of the papers reviewed dealt with forecasting incoming call volume traffic in a call center using historical data. The majority of papers using statistical models are linear models that will not capture the non-linearity dependence of the incoming call volume traffic. However, the hybrid model (statistical and Neural network) improves forecast performance, as shown in [16]. However, these researches have some gaps that do not incorporate all the characteristics of incoming call traffic volumes, i.e., long term dependency and complexity.

1.8 Thesis Organization

The following is how the remainder of this thesis is organized: Chapter Two discusses the call center overview and its basic architectural components. The time series call volume modeling and forecasting techniques is discussed in Chapter Three. The fourth chapter describes the results and discussion of an experiment. Finally, Chapter Five discusses the conclusion and recommendation.

CHAPTER TWO

Overview of Call Center

In this section, an overview of call center service is presented followed by the general call center architecture and definition of its components.

2.1 Overview

A call center is a department or office where a team of advisors can handle the incoming and outgoing phone calls from both new and existing consumers.

A call center is a centralized office dedicated to receiving and transmitting enormous amounts of data by telephone and computer. Larger companies require call centers to sell or market their products and services, as well as to provide after-sales services and answer client questions. The call center is another tool for customer relationship management that organizations can utilize to enhance customer experience and increase efficiency. It serves as an interface between customers and businesses, allowing them to call and report problems or inquire about products or services. The call center manager role is to assist the call center workers by resolving customer concerns that they are unable to resolve [18].

The call center can be a profit center, a cost center, a major source of revenue, a major source of customer dissatisfaction, a strategic disadvantage, a source of marketing research, and so on. The value of call center services varies by company, depending on how closely the call center collaborates with parent or customer organizations to support the company. Call centers are the "voice" of the company for customers. Depending on the type of call center, there are several ways to communicate with customers[12].

Call centers are classified into two types: inbound and outbound. Customers initiate calls to inbound call centers. Customer service, technical support for business or private customers, downloading, entering, and processing orders, scheduling appointments, reservations, and registration are all services provided by inbound call centers. Agents from the company initiate calls to customers in an outbound call center. [3].

Some of the most important Call Center management features are[12]:

- Allows incoming phone calls to be properly categorized and qualified without using live agent resources;
- Responds quickly to customer calls;
- Solves a high percentage of customer queries on the first call, which has good results in terms of customer satisfaction;

- Allows monitoring phone calls, which plays an important role in improving and optimizing call center efficiency;

2.2 General Call Center Architecture

The most effective call center layouts designs are those that increase productivity while reducing the customer dissatisfaction. Showing up on time, being willing to learn, and having a generally positive attitude are all important traits for call center representatives. Empathy, exceptional listening skills (and a willingness to listen to the customer) are also important traits for call center representatives [19]. The Figure 2.1 shows the general call center architecture.

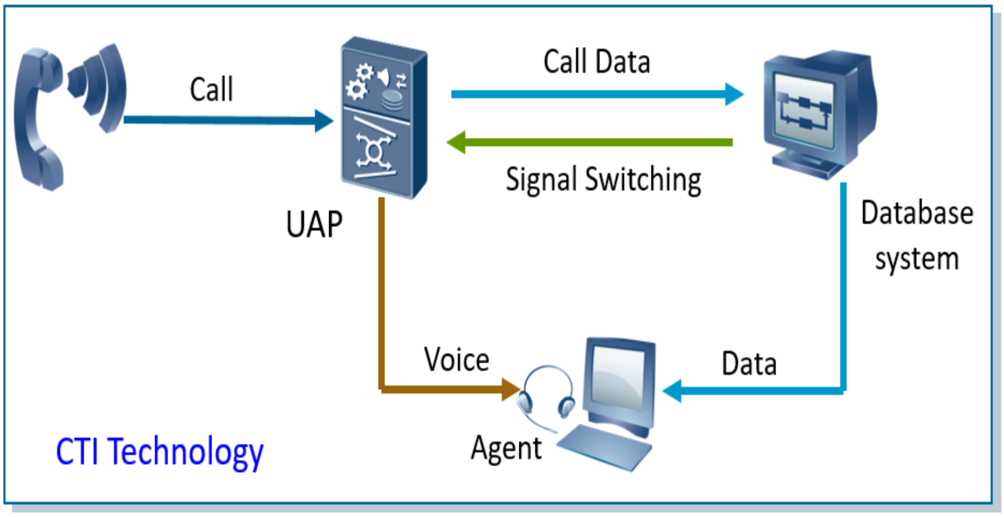


Figure 2. 1 general call center *architecture*

Ethio Telecom uses Huawei IPCC system in call center service for facilitate service. Huawei IP Contact Center (IPCC) system is an intelligent device platform with personalized processing capability. It integrates the access processing technologies of media such as the Internet, VoIP, E-mail, fax and video. In IPCC system there are two important platforms: UAP (Universal Access Platform) and CTI platform (Computer & Telephony Integration)

- **CTI Technology**: It is a technology that allows interactions on a telephone and a computer to be integrated or coordinated.

- **UAP**: The Universal Access Platform is an integrated broadband and narrowband access platform developed by HUAWEI. The UAP has powerful networking capability. It supports the functions such as call access, agents, and media resources for a call center.

The following components describe the generic technology that enables the integration of a call center:

- **Telephone:** Activated by an integrated Call Center that employs user-friendly technology such as telephone switches, automatic call distributors, voice processing, computer telephony interfaces, and other customer care applications.
- **Self-Service:** Enabled by incorporating Interactive Voice Response (IVR), which allows the public to either surf the agency's portal or call to obtain the necessary information through a series of clicks or key presses.
- **Interactive Voice Response (IVR):** Allows customers to first communicate their needs to the IVR. The technology will be able to determine the most appropriate agent to handle the call based on the customers' requirements.
- **Agents:** Customers are attended to by agents, sometimes known as Customer Service Officers. They can access consumer requests, calls, and complaints through a variety of customer contact points. They can also access the different back-end systems that help them run their business.
- **Business Applications:** Business applications are software programs that assist agents in their regular tasks. These technologies allow agents to collect vital information about their clients. This can come from Customer Relationship Management (CRM) systems or even legacy back office systems.
- **Operational Data Stores & Warehouses:** These are areas where customer information is stored. These data stores will be used by business applications to obtain customer information. While there may be a variety of business applications and data storages on individual customers, it is critical that all of the business applications have a single view of the customer.
- **Automatic Call Distributor (ACD):** can be combined with intelligent routers to facilitate virtual Call Centers by analyzing and distributing large volumes of calls to agents/CSOs, as well as capturing and storing extensive information relevant to those calls – queue to queue transfer, group and personal call backs, and ACD can be combined with intelligent routers to facilitate virtual Call Centers.

CHAPTER THREE

Time Series Call Volume Modeling and Forecasting Techniques

This chapter discusses various types of time series forecasting models, as well as their components. Particularly, it focuses on statistical forecasting models and deep learning models. These models are described below in detail.

3.1 Definition of Time Series

A time series is a sequence of data points in which the orders indicate the sequence of measurable over time. A time series is defined mathematically by the values x_1, x_2, x_3,...x_n of the variable x at times t_1, t_2, t_2,...t_n. A time series may be continuous or discrete. A continuous time series includes observations taken at every point in time, whereas a discrete time series includes observations taken at discrete points in time. A univariate time series is one that only contains observation data for a single variable. However, when records from more than one variable are considered, it is referred to as multivariate [20].

3.2 Time Series Analysis

Time series analysis involves methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series are studied for a variety of purposes such as the forecasting of the future based on information of the past, understanding of the phenomenon underlying the measures, or simply a concise description of the significant features of the series [21].

3.2.1 Components of Time Series

Time series data contains a variety of patterns, and it is often helpful to split a time series into several components, each representing a different underlying pattern category. Often this is done to help improve understanding of the time series, but it can also be used to improve forecast accuracy. A time series can be decomposed into four major components: *trend*, *seasonal*, *cyclical*, and *residual*; each of them are described as follows [22].

3.2.1.1 Trend Components

The trend is a non-repeating and long term pattern of a time series. A trend can be either positive or negative depending on whether the time series shows an upward long term pattern or a downward long term pattern.

3.2.1.2 Seasonality Components

Seasonality describes a repeating behavior that occurs over the short term at the predictable interval, which spans less than a year, such as hourly, weekly, monthly, or quarter.

3.2.1.3 Residual Components

The residual component is unpredictable. Every time series has some unpredictable component that makes it a random variable. In prediction, the objective is to "model" all the components to the point that the only component that remains unexplained is the random component.

3.2.1.4 Cyclic Components

A cyclical pattern is defined as any pattern that shows an up and down movement around a given trend. The length of a cycle is determined by the type of business or industry being studied.

For decomposed time series data, two types of mathematical models are typically used to account for the effects of trend, seasonality, residual and cyclic components. These models are additive and multiplicative models. The additive model works based on the assumption that the four components are independent of one another, whereas the multiplicative model works based on the assumption that the four components of a time series are dependent on one another. Mathematically representation of the additive and the multiplicative models are described in Equations (2.1) and (2.2) respectively.

- Additive model equation

$$Y(t) = T(t) + S(t) + C(t) + R(t) \qquad (2.1)$$

- Multiplicative model equation

$$Y(t) = T(t) \times S(t) \times C(t) \times R(t) \qquad (2.2)$$

Where Y(t) is the observation and T(t) + S(t) + C(t) + R(t) are the trend, seasonal, cyclical, and residual at time t respectively.

3.3 Time Series Forecasting

A time series forecasting method is a technique for predicting the future value based on current and historical data. In general time series forecasting techniques are classified into a linear model, nonlinear model, hybrid model and decomposed model[20].

- **Linear time series models:** linear time series model is to study the dynamic structure of such a series data. The two main subgroups of this technique are the Auto Regressive (AR) and Moving Average (MA) models. The auto regressive moving average model (ARMA) and seasonal auto regressive integrated moving average (SARIMA) model are created by combining these two models.
- **Nonlinear time series model:** are used to investigate aspects that linear processes can't handle, such as cycle and time-change variance. Techniques such as Neural Networks are one example[23].
- **Hybrid model:** It is mostly made up of a mix of linear and nonlinear models. For instance, ARIMA with an Artificial Neural Network.
- **Decomposed model:** is implemented by splitting down the time series into seasonal, trend, cyclical, and irregular components. For example, the nonlinear decomposed model decomposes time series into trend, period, mutation, and random components.

There are different time series forecasting methods used for various applications such as ARIMA, Holt-Winters smoothing, SARIMA, and seasonal moving average models in traditional forecasting methods [12] [13] and [16]. Another modern deep learning time series forecasting methods are multilayer perceptron (MLP) and LSTM. The statistical forecasting methods handle only linear data behavior. Whereas, the deep learning algorithms are used for time series forecasting methods to handle non-linear data behavior. From the traditional methods, SARIMA is the most common method used for time series forecasting applications that is proposed in this thesis work. On the other hand, LSTM is proposed and used in this work because it captures long term dependency and complexity that can also increase the accuracy of the model. Details of these models are discussed in the subsequent sections.

3.4 Statistical Forecasting Model

A statistical model is a mathematical representation of observation data that attempts to analytically determine the relationship between two or more random variables. Forecasting time series models are either linear or non-linear, depending on whether the current value of the series is a linear or non-linear function of past observations. In general time series data models can take many forms and represent various stochastic processes [21]. Some examples of statistical forecasting models include the autoregressive (AR), moving average (MA), and ARIMA models. Box and Jenkins [24] proposed a very successful variation of the ARIMA model for seasonal time series forecasting, namely the Seasonal ARIMA (SARIMA) model [22][13][25][26].

In this section we discuss some of the linear statistical models with mathematical expressions.

3.4.1 AR Model

AR model forecasts future variables based on past values. AR models are based on the idea that the current value of the series, X_t, can be explained as a linear combination of p past values, X_{t-1} $+X_{t-2}+...X_{t-p}+W_t$ where W_t is White noise in the same series. AR (p) is a common notation for an AR model, where p is the model order. The term "auto-regression" refers to a regression of the variable itself. The AR model's order indicates how many lagged past values are included. The mathematical description is shown in Equation (2.3).

$$x(t) = \sum_{j=1}^{P} a_j x(t-j) + w(t) \qquad (2.3)$$

Where,

- $x(t-j)$ is the previous values sample;
- a_j is auto regressive coefficients at order of p;
- $w(t)$ is White noise with zero mean.

3.4.2 Moving Average Model

The MA model, rather than using past forecast values in a regression, uses past forecast errors in a regression-like model. The MA model considers a sample process to be a moving average (unequally weighted) of a random sample process. The moving average model's order is denoted by the letter q. The common notation of MA is MA(q). The mathematical description is shown in Equation (2.4).

$$x(t) = \sum_{j=1}^{q} b_j w(t-j) + w(t) \tag{2.4}$$

Where,

- $w(t-j)$ is the previous values sample;
- b_j is auto regressive coefficients at order of p;
- $w(t)$ is white noise with zero mean.

3.4.3 ARMA Model

ARMA model is obtained from two models (a hybrid of an autoregressive and a moving average model). In time series analysis, the ARMA model is used to describe stationary time series. The ARMA model represents time series produced by sequentially passing white noise through a recursive and a non-recursive linear filter. The common nation of ARMA model is ARMA (p,q); where:

- p is the order of the autoregressive polynomial,
- q is the order of the moving average polynomial.

The ARMA model of mathematical description is shown in Equation (2.5).

$$x(t) = w(t) + \sum_{j=1}^{q} a_j x(t-j) + \sum_{j=1}^{q} b_j w(t-j) \tag{2.5}$$

3.4.4 ARIMA Model

The ARIMA model is one of the most widely used and well-known stochastic time series models. The integrated ARMA (ARIMA) is a type of ARMA that includes differencing [21]. ARIMA model is the most general class of models for forecasting a time series which can be made to be "stationary" by differencing (if necessary), perhaps in conjunction with nonlinear transformations such as logging or deflating (if necessary). The ARIMA model can be viewed as a "filter" that tries to separate the signal from the noise, and the signal is then extrapolated into the future to obtain forecasts. The ARMA procedure offers a comprehensive set of tools for univariate time series model identification, parameter estimation, and forecasting, as well as a high level of flexibility in the types of ARIMA models that can be analyzed. Seasonal, subset, and factored ARIMA models are supported, as are intervention or interrupted time series models, multiple regression analysis with ARMA errors, and rational transfer function models of any complexity [24].

The ARIMA model is an ARMA model that is based on a differenced series, which is created by combining differencing with AR and MA models. The ARIMA model is denoted as the ARIMA (p, d, q) model, where p represents the number of autoregressive terms, d represents the number of non-seasonal differences required for stationary, and q represents the number of moving average terms. The mathematical description of the ARIMA model is shown in Equation (2.6).

$$x(t) = c + \sum_{j=1}^{q} a_j x(t-j) + \sum_{j=1}^{q} b_j w(t-j) + w(t) \qquad (2.6)$$

3.4.5 SARIMA Model

SARIMA models are an extension of ARIMA that explicitly supports univariate time series data with seasonal components. This model is divided into two parts: non-seasonal and seasonal, with the AR, Integrated, and MA parameters in each. It is used when there is a periodic characteristic in the data that must be known ahead of time. SARIMA (p,d,q)(P,D,Q)s is the general form of a seasonal ARIMA model, where p is the non-seasonal AR order, d is the non-seasonal differencing, q is the non-seasonal MA order, P is the seasonal AR order, D is the seasonal differencing, Q is the seasonal MA order, and S is the time steps of repeating seasonal pattern.

3.4.5.1 SARIMA Modeling Steps

The most crucial step in estimating the SARIMA model is determining the values of seven parameters (p,d,q)(P,D,Q,s). If the variance grows with time, for example, based on the time plot of the data, we should use variance-stabilizing transformations and differences. Then, using the ACF to determine how much linear dependence exists between observations in a time series separated by a lag p and the PACF to determine how many autoregressive terms q is required, we can identify the preliminary values of autoregressive order p, order of differencing d, moving average order q, and their corresponding seasonal parameters P, D, and Q.

1. **Stationary process**

Many time series techniques make the assumption that the data are stationary. A stationary process has the statistical property that the mean, variance, and autocorrelation structure do not change over time[22]. The requirements for achieving stationarity necessitate the fulfillment of the three constraints listed below:

- The mean value has to be approximated to a constant: $E(X_t)$= constant for all t;
- The variance has to be approximated to a constant: Variance (X_t)= constant for all t;
- The co-variance must be dependent on lag j: $Cov(X_{t,} T_{t+j})$ =constant for all t.

2. **Augmented Dickey Fuller Test**

The Augmented Dickey Fuller test (ADF Test) is a unit root test that used to determine whether or not a given time series is stationary. It is one of the most commonly used statistical tests for analyzing the stationary of a series. To perform an Augmented Dickey-Fuller test in python it returns the p-value.

The null hypothesis of the ADF test is that the time series is non-stationary. So, if the p-value of the test is less than the significance level (0.05) then reject the null hypothesis and infer that the time series is indeed stationary. So, in our case, if P Value > 0.05, we go ahead with finding the order of differencing. When the P Value > 0.05, the null hypothesis is rejected and the time series is assumed to be stationary. The null hypothesis is based on the assumption that the time series is non-stationary.

3. **Autocorrelation Function and Partial Autocorrelation Function plot**

The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots graphically summarize the strength of a relationship between an observation in a time series and observations at previous time steps. To investigate the linear relationship between two variables, the ACF is used. The ACF is a method for calculating the linear relationship between an observation at time t (the current time) and previous observations. The PACF summarizes the relationship between an observation in a time series and observations at previous time steps by removing the relationships of intervening observations.

4. **Akaike Information Criterion**

The Akaike Information Criterion (AIC) is a mathematical method for assessing how well a model fits the data from which it was generated. AIC is used in statistics to compare different possible models and determine which one is the best fit for the data. The AIC is calculated using the number of predictor variables used to build the model and the maximum likelihood estimate of the model (how well the model reproduces the data).

3.5 Deep Learning Forecasting Model

Deep learning (DL) algorithms are a more advanced and mathematically complex evolution of machine learning algorithms. DL is a subfield of machine learning, with neural networks serving as the foundation of deep learning algorithms. It analyzes data in the same way that humans do. DL methods, such as automatic learning of temporal dependence and automatic handling of temporal structures such as trends and seasonality, hold a lot of promise for time series forecasting. DL models for time series forecasting come in a variety of flavors such as Artificial Neural Network (ANN) model, Recurrent Neural Network (RNN) model and LSTM model.

3.5.1 Artificial Neural Network Model

The ANN model is a type of intelligent system that can be used to solve complex problems in a variety of applications, such as optimization, prediction, modeling, clustering, pattern recognition, simulation, and more [27][28]. Three layers of artificial neurons or nodes make up the ANN structure: an input layer that collects data, an output layer that computes information, and one or more hidden layers that connect the input and output layers. Each of the ANN layers contains neurons, which are nodes. A layer is a collection of nodes that have the same input and output connections but don't communicate with one another in the same layer. Figure 3.1 shows the ANN general architecture with three layers.

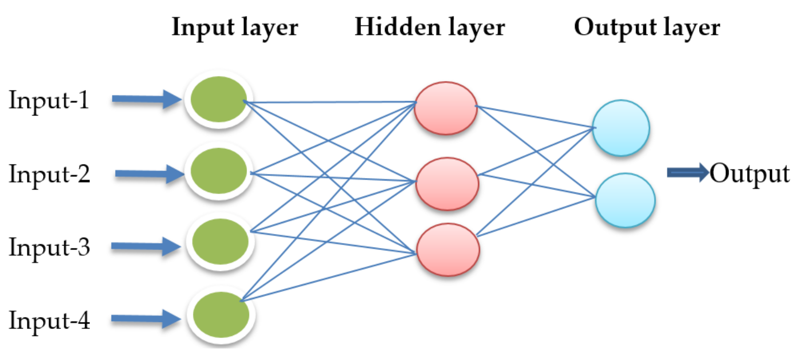


Figure 3. 1 ANN architecture with 3 layers [28]

A neuron is a mathematical function that simulates the operation of a biological neuron within an artificial neural network. A neuron typically computes the weighted average of its input and then passes this sum through an activation function. The number of features in the dataset determines the dimensionality, or the number of nodes, in the input layer. "Synapses" connect these nodes to the nodes created in the hidden layers. The synapses links have some weights for each node in the input layer. The weights function as a decision maker, determining which signals or inputs may or may not pass through. The weights also represent the strength and depth of the hidden layer. A neural network learns by varying the weight for each synopsis.

3.5.2 Recurrent Neural Network Model

RNN is a class of artificial neural networks [29] that is commonly used in natural language processing and time series forecasting. The RNN model employs sequential observations, with the goal of predicting the next step in the sequence of observations based on the previous steps observed in the sequence. The RNN model adds a hidden state that is generated by the sequential information of a time series, with the output dependent on the hidden state. Figure 3.2 shows the most common architecture of the RNN model begins to unfold into a fully connected network.

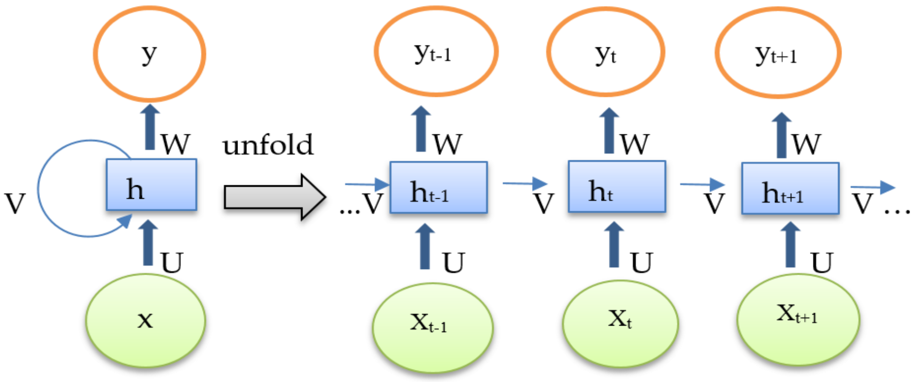


Figure 3. 2 RNN model architecture [29]

The mathematical expression for the RNN model in Figure 3.2 is denoted in Equations (2.7) and (2.8):

$$h_t = f(Ux_t + Vh_{t-1}) \quad (2.7)$$

$$y_t = f(Wh_t) \quad (2.8)$$

Where;

- X_t is the input at time t;
- U denotes weight matrix from the input layer to the hidden layer;
- V denotes the weight of recurrent computation;
- W denotes weight from hidden layer to the output layer;
- h_t denotes values of hidden nodes at the time of t;
- y_t denotes a value of output node at the time of t;
- f is the activate function, which has many alternatives such as sigmoid function and ReLU.

The RNN model is well modeled time series data that is the temporal dependency [29]. The main problem with a typical generic RNN is that these networks remember only a short term dependence in the sequence. So it is hard to remember longer dependency within sequences of data due to the vanishing gradients problem. This problem is solved by utilizing the "memory" introduced in the LSTM recurrent network [30].

3.5.3 LSTM Model

The LSTM model is a type of recurrent neural network that can remember patterns selectively for long periods of time. The hidden layer unit in the original RNN architecture is replaced by a memory blocks (called cells) in the LSTM architecture. It carries all the information with only some linear interaction. The LSTM does have the ability to remove or add information to the cell state, carefully regulated by structures called gates. The gates, which are based on a sigmoid neural network layer, allow cells to either pass data through or discard it. Each sigmoid layer produces a number between 0 and 1, indicating how much of each data segment should be allowed through in each cell. A value of zero means "let nothing through," while a value of one means "let everything through!" [31]. Figure 3.3 shows the LSTM model architecture.

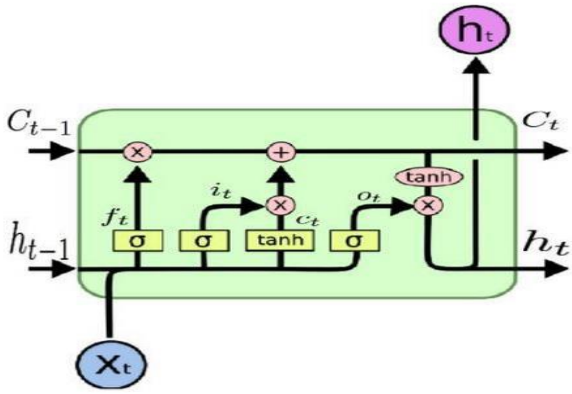


Figure 3. 3 LSTM model architecture [31]

LSTM networks are typically made up of memory blocks known as cells that are linked together via layers. Cell state Ct and hidden state h_t contain information, which is regulated by mechanisms known as gates via sigmoid and tanh activation functions. As a result, LSTM can conditionally add or delete information from the cell state. In general, the gates take the hidden states from the previous time step h_{t-1} and the current input x_t as inputs and multiply them point wise by weight matrices with a bias added to the product.

The LSTM network unit consists of a cell, an input gate, an output gate, and a forget gate. A cell retains values for an autocratic time interval. The input gate controls the flow of information into the cell. The output gate controls the flow of information to and from the outside world. Similarly, forget gates control the flow of information that is necessary or unnecessary. All of the LSTM network units is provided in detail below:

A. Forget gate

It aids in determining whether information can pass through the network's layers. It expects two types of input from the network: information from previous layers and information from the presentation layer. In general, the forget gate layer decides what information from the previous cell state to retain. The mathematics is represented in Equation (2.9).

$$f_t = \sigma(w_f * [h_{t-1}, x_t] + b_f) \tag{2.9}$$

B. Input gate

The following step is to decide what new information we will store in the cell state. This is divided into two parts. First, a sigmoid layer known as the "input gate layer" determines which values to update. A tanh layer then generates a vector of new candidate values, ct, that could be added to the state. In the following step, we'll combine these two to create a state update. The information that must be stored is then determined by the following two Equations.

$$i_t = \sigma(w_i * [h_{t-1}, x_t] + b_i) \tag{2.10}$$

$$c_t = \mathrm{tnah}(w_c * [h_{t-1}, x_t] + b_c) \tag{2.11}$$

C. Output gate

Finally, the output gate is the last gate that helps decide how much information from the current cell state flows into the hidden state. This output will be based on our cell state, but will be a filtered version. First, we run a sigmoid layer which decides what parts of the cell state we're going to output. Then, we put the cell state through tanh (to push the values to be between −1 and 1) and multiply it by the output of the sigmoid gate, so that we only output the parts we decided to. The mathematical representation in Equations (2.12) and (2.13).

$$o_t = \sigma(w_o * [h_{t-1}, x_t] + b_o) \tag{2.12}$$

$$h_t = \mathrm{tnah}(C_t) \tag{2.13}$$

In the literature, several statistical models have been proposed to forecast time series for incoming call volume traffic in the context of call center service. ARIMA, Holt-Winters smoothing, and SARIMA models are the most widely used traditional forecasting methods [12] [13] and [16] and [26]. The data generating function is constrained by the inherent constraint of linearity in these models. To address this, a number of nonlinear models have been developed in the literature.

Deep learning algorithms, in particular, have introduced new approaches to prediction problems in which the relationships between variables are modeled in a deep and layered hierarchy. Deep learning-based algorithms like RNN and LSTM have gotten a lot of attention in recent years due to their applications in a variety of fields, including finance. Deep learning methods can detect data structure and pattern, such as nonlinearity and complexity in time series forecasting [31].

The main goal of this thesis is to compare statistical and deep neural network models for forecasting incoming call volumes, using the SARIMA and LSTM algorithms. Which forecasting methods provide the best predictions in terms of lower forecast errors and higher accuracy of forecasts. Using these proposed methods, we'll do our experiment and obtain the required results in the next chapter.

CHAPTER FOUR

Experimental and Result Analysis

In this chapter, we describe procedures followed to conduct the experiment and results obtained. The system model building step is described in Section 4.1. After that Sections 4.2 and 4.3 discuss Data collection and Data preprocessing respectively. Whereas Section 4.4 explains the Algorithms used in this study, then Section 4.5 is about Performance Metrics. Finally, Results and Discussion are presented in Section 4.6.

4.1 System Model

The system model presents the steps to be followed for forecasting incoming call volume in the call center. Figure 4.1 represents the system model to perform modeling of incoming call volumes forecast.

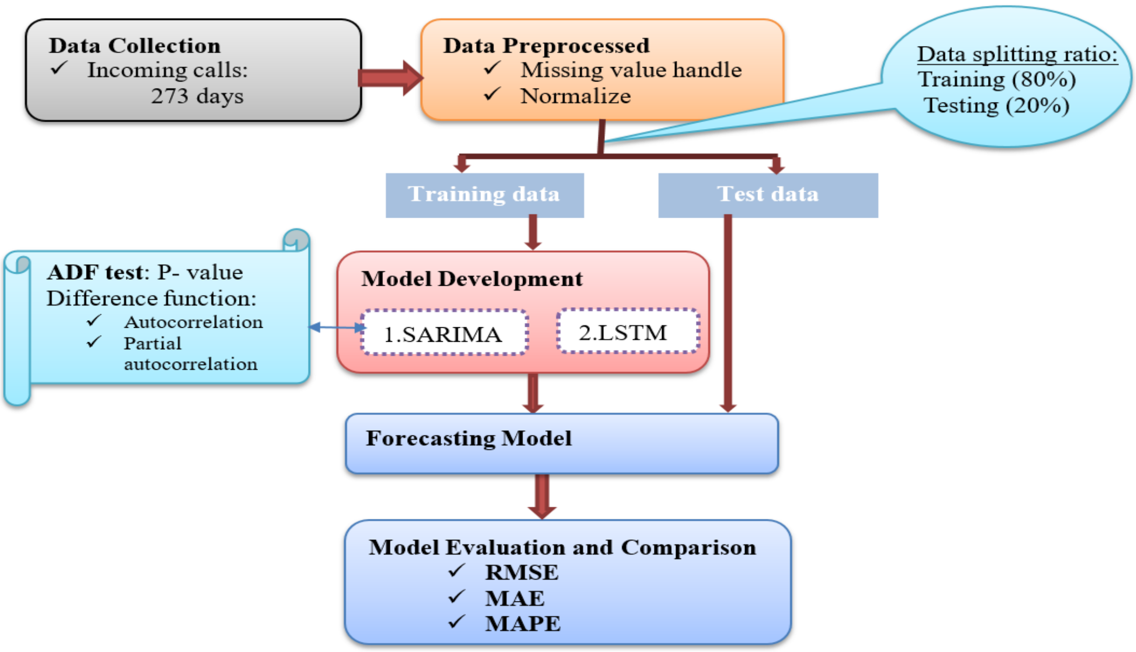


Figure 4. 1 System Model

System models are described as follows:

- System model begins with data collection.

- Then the collected data is pre-processed. Firstly, data is visualized, and handling missing values and data normalization are performed. Then, the dataset is divided into two groups: training set (80%) and testing set (20%).
- Then next SARIMA and LSTM algorithms are trained using the 80% training dataset.
- The SARIMA and LSTM models are then tested using the 20% testing dataset.
- The models are evaluated based on different performance metrics such as RMSE, MAE and MAPE.
- Finally, the one with the minimum prediction error is selected and recommended.

4.2 Data Collection

The data are gathered at the inbound call center operations IPCC system of Ethio telecom. It is collected for over 273 days from August 1, 2020 to April 30, 2021, on a half hourly basis. The call center operates 16 hours per day from 7:00AM to 11:00, PM seven days a week. In this study, a time series of incoming call volume data is considered for understanding the incoming call volume pattern and forecasting future incoming call volume.

4.3 Data Preprocessing

During the data preprocessing task, raw data can be reshaped to an understandable form. The data is collected daily on a half-hourly basis from the call center IPCC system. Incoming call volumes traffic accounted for 273 days of total data usage. The call arrival counts are aggregated in thirty-minute length, with each day's data covering the same time period. The total number of recorded observations is 32*273 = 8,736 data set. In the data analysis, MS Excel and the Python tools are used. There are 3 hours missing values in the data, one hour missing value on August 20, 2020, and the other two hour January 28, 2021. The missing values are filled using a linear interpolation method.

4.3.1 Data Visualization

Time series plots of raw sample data can provide valuable diagnostics for identifying temporal structures such as trends, cycles, and seasonality that can influence model selection. Figure 4.1 shows time series plot daily incoming call volumes in Ethio Telecom call center from August 1, 2020 – April 30, 2021.

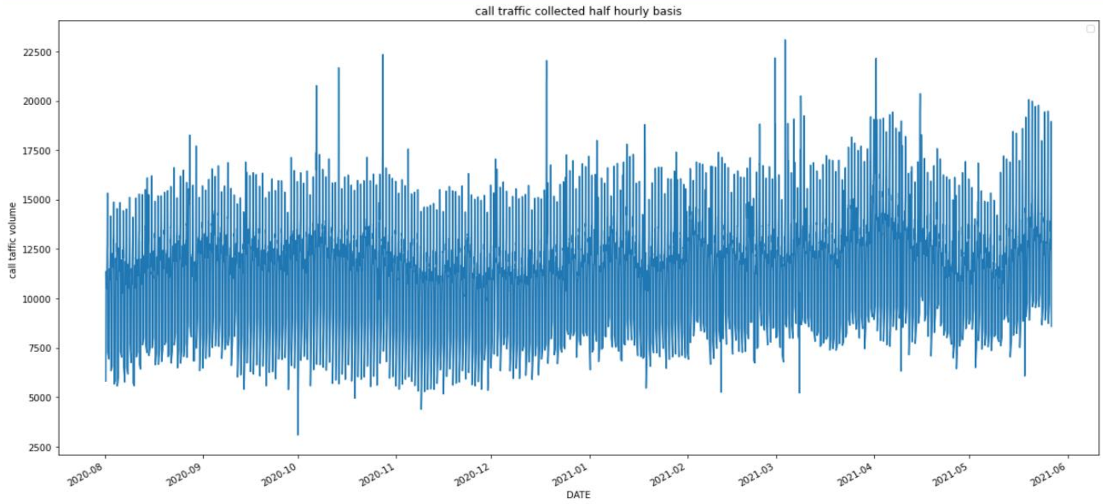


Figure 4. 2 Time series of half-hourly basis incoming call volumes in Ethio telecom inbound call center

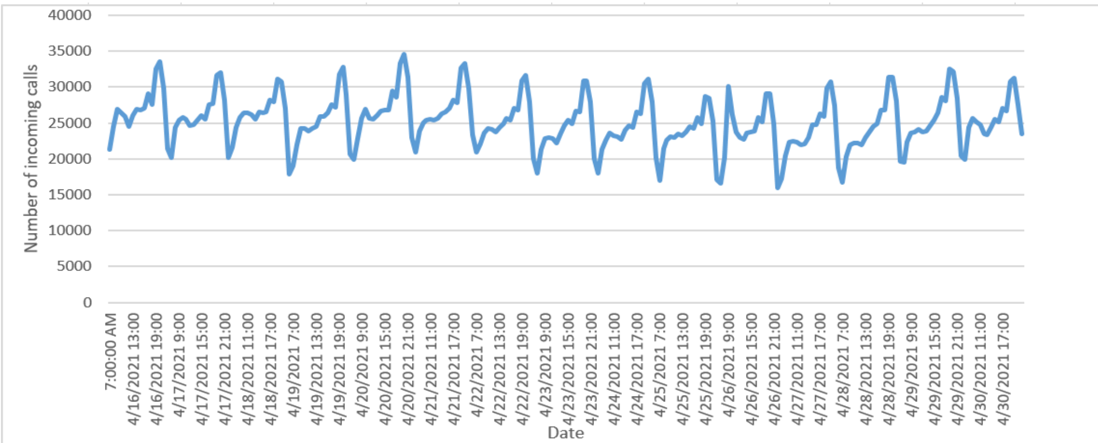


Figure 4. 3 Number of incoming call volumes for two weeks from April 16 to 30,2021

Figure 4.3 shows the number of incoming call volume patterns for two weeks of the Ethio Telecom call center. The number of incoming calls follows a similar usage pattern each day for the two weeks. There are low and high peaks in incoming call volumes during the day. The first low peak occurs in the morning, from 7:00 AM up to 11:00 AM and the second low peak occurs in the night from 21:00 PM up to 23:00 PM. The first high peak occurs in the morning, from 11:00 AM up to afternoon 16:00 PM and the second, high peak call volume occurs in the afternoon from 16:00 PM up to night 21:00 PM.

4.3.2 Training Dataset and Test Dataset

We have used a total dataset of 8736 incoming call volume traffic. This dataset is divided into two parts: training data and test data. The training data set is used to fit (train) a model for prediction, whereas the test data set is used to evaluate model prediction accuracy. The data split ratio is defined as the first 80% of the total data set for training and the remaining 20% for testing.

4.3.3 Time Series Decomposition

Decomposition is primarily used for time series analysis, but it can also be used as an analysis tool to inform forecasting models on your problem. Figure 4.4 shows the original time series data decomposition into components (trend, seasonal and residual) of the incoming call volumes traffic. It examines features such as up and down trends and seasonality. The pattern across those time units shows a daily seasonal pattern. The observed regularly repeating pattern of highs and lows relates to seasonal data of each day, which show seasonality.

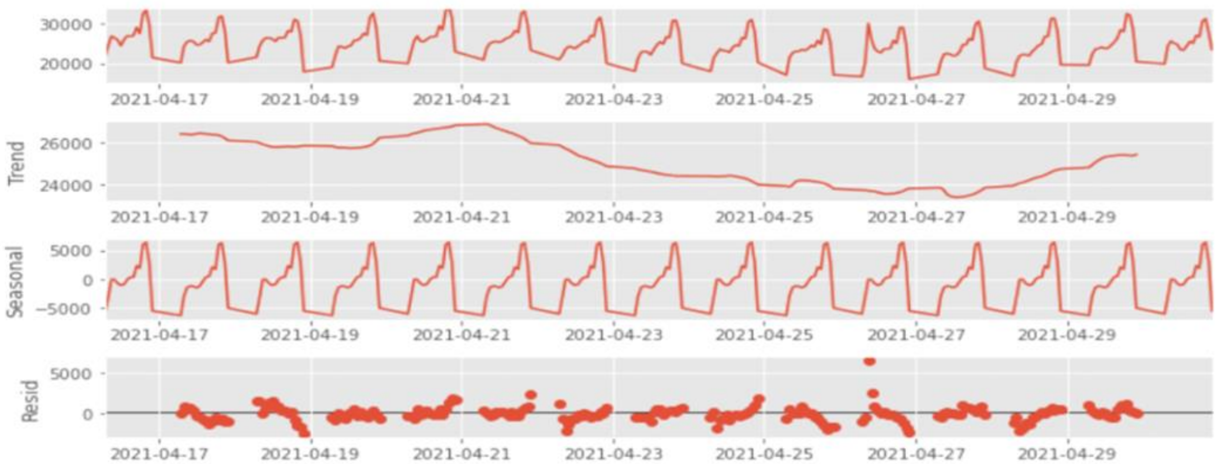


Figure 4. 4 Time series of incoming calls decomposed into trend, seasonality and residuals

4.4 Algorithms for Call Volume Forecasting

4.4.1 The SARIMA Model

SARIMA models are an extension of ARIMA that explicitly supports univariate time series data with a seasonal component. The most crucial step in estimating the SARIMA model is determining the values of model seven parameters. To determine the value of SARIMA model parameters using ACF and PACF plots and grid search. The ACF and PACF plots were used as a starting point to help determine a few likely parameters, and then the best parameters were identified using a grid search.

4.4.1.1 SARIMA Model Parameter Analysis

1. ACF and PACF Plot

Plotting the ACF and PACF graphs of time series data is critical for identifying all of the initial SARIMA model parameter values of p, d, q, and P, D, Q. Figure 4.5 depicts the ACF and PACF plots of the original time series prior to data transfer. The ACF graph shows a high peak that repeats every 32 lags and a positive spike in lags 1 and 2. In the PACF graph, there is a positive and negative spike in lags 1 and 2 and every 32 lags repeats.

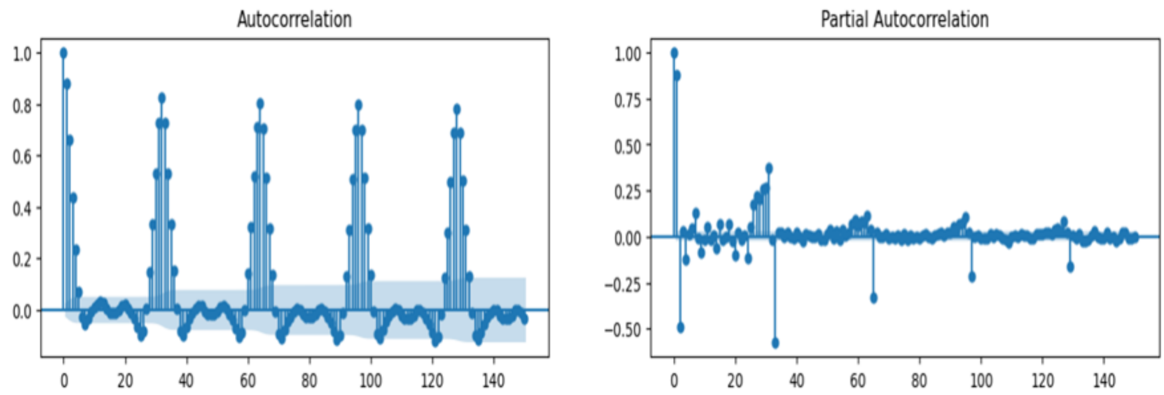


Figure 4. 5 ACF and PACF plot, x axis represents number of lag and y axis represents correlation

It's clear from the figure that the data contains a seasonality component, so transformations will need to be applied to make the data stationary. To make the data stationary it applied the log transform and difference functions. Figure 4.6 shows the original data have been transformed and differenced using the log transform and difference functions the time series data and ACF and PACF plots.

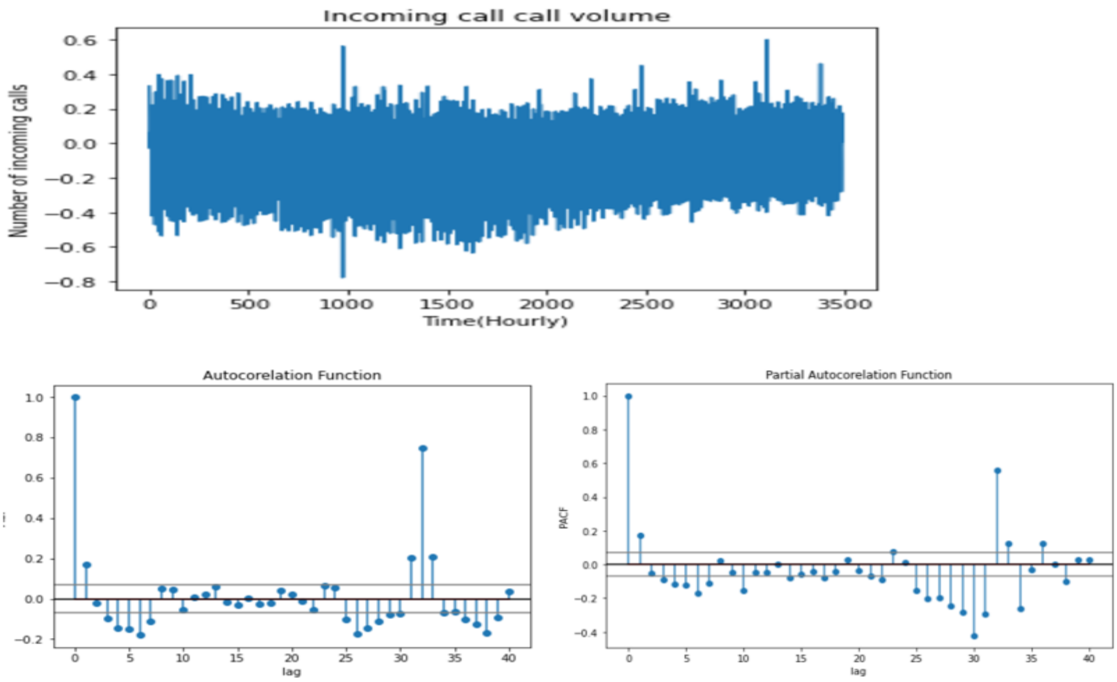


Figure 4. 6 Time series plot after log transformed and differenced, ACF and PACF plot

2. Stationary Test

Test the stationary of the time series through a unit root test, ADF test and examine its incoming call volume traffic for trend and seasonality. A statistically significant test results the null hypothesis (p-value > 0.05) for this test is that the data is not stationary and the null hypothesis (p-value < 0.05) for this test is that the data is stationary. The statistical test result from the ADF test the p-value was 0.022286588599246416. Based on the result the p-value <0.05, so the data is stationary.

3. Model Selection

The SARIMA algorithm is applied to forecast the incoming call volume data. The grid search method is used for hyperparameter optimization and determining the appropriate forecasting model parameters. The initial value of SARIMA model parameters is determined through ACF and PACF plots, SARIMA (1, 1, 1) (1, 1, 1)32. With a starting point of (1, 1, 1) (1, 1, 1)32 a grid search was set up to test several different parameter combinations. The models were evaluated using the AIC criterion. Table 4.2 displays some of the model grid search results that show the models with the lowest AIC score values. So the final SARIMA model parameters are used in SARIMA (1, 1, 1) (1, 0, 1)32.

Non seasonal parameter	Seasonal parameter	Model selection criterion
p,d,q	P,D,Q, s	AIC value
0, 0, 0	0, 0, 0, 32	100707.53049956143
0, 0, 0	0, 0, 1, 32	95263.44944524925
0, 0, 1	1, 0, 1, 32	79196.9910507604
0, 0, 1	1, 1, 0, 32	74248.0083825425
0, 0, 1	0, 1, 1, 32	73284.11611560354
0, 0, 1	1, 0, 0, 32	79220.64498605275
0, 1, 0	0, 0, 0, 32	77969.06021136619
0, 1, 0	0, 0, 1, 32	75541.24142485256
0, 1, 0	0, 1, 0, 32	72620.21327138646
1, 1, 0	1, 0, 0, 32	72660.58794614015
1, 1, 0	1, 0, 1, 32	71725.05428571094
1, 1, 1	1, 1, 1, 32	69927.94363873173
1, 1, 1	1, 0, 1, 32	10.00

Table 4. 1 AIC values of SARIMA models

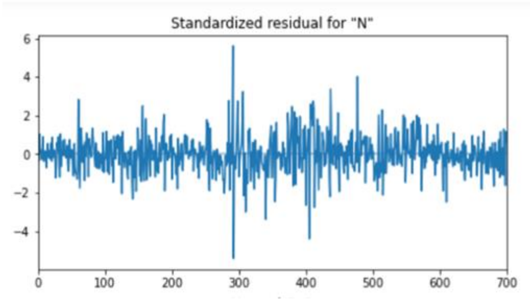
4. Diagnosis check

A diagnosis check is used to validate the selected model's forecast accuracy. The AIC value of SARIMA (1, 1, 1) (1, 0, 1)32 is the lowest as shown in Table 4.1. Table 4.2 depicts the diagnostic check of the chosen SARIMA model summary, which shows that the p-value is less than the significant value (0.05), so this indicates that the forecast accuracy of the chosen model is well. In the table also shows the SARIMA parameter value of lags.

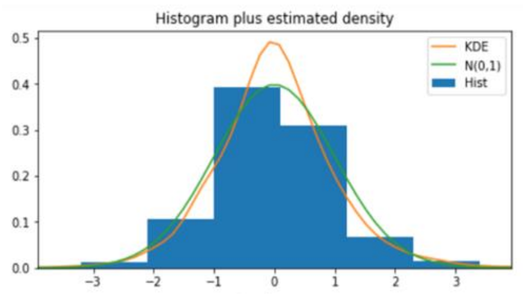
```
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8642      0.003    302.241      0.000       0.859       0.870
ma.L1         -1.0000      0.199     -5.029      0.000      -1.390      -0.610
ma.S.L32      -0.8675      0.004   -216.800      0.000      -0.875      -0.860
sigma2         0.0016      0.000      5.000      0.000       0.001       0.002
==============================================================================
```

Table 4. 2 The diagnostics test results of the SARIMA (1, 1, 1) (1, 0, 1,)$_{32}$ model.

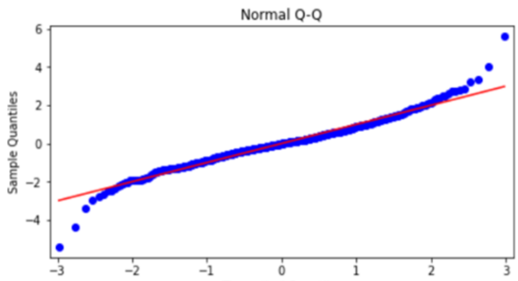
5. Residual Plot



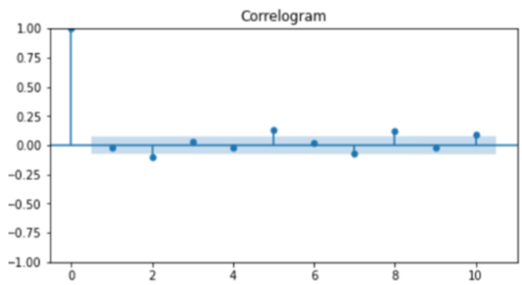
(a)



(b)



(c)



(d)

Figure 4. 7 Plot of residual: (a) Residuals over time; (b) Distribution histogram; (c) Q-Q plot and (d) Autocorrelation

Figure 4.8, (a), depicts the residuals over time. The findings imply that the residuals have no discernible seasonality and appear to be white noise. Similarly, the autocorrelation in Figure 4.8 (d) indicates that the residuals of the original data have a low correlation with the lagged data. According to Figure 4.8, (b), the Kernel Density Estimation (KDE) (red curve) is nearly overlapping with the N (0, 1) (green curve). The results indicate that the residual has a normal distribution, with a mean of 0 and a standard deviation of 1. The red line in Figure 4.8 (c) represents normally distributed incoming call volumes with a mean of zero and a standard deviation of one, while the blue dots represent residuals. In general, the Q-Q plot shows residuals that are normally distributed, indicating that the chosen model fits well and can be used to forecast future values.

4.4.2 LSTM Model

The LSTM is a RNN that can store and learn from long sequences of observations. The developed LSTM model is a multi-step univariate forecast algorithm. We set the time steps 32 data point past values used to predict future one week ahead (from one half hour up to five days) incoming call volumes traffic. The incoming call volume traffic prediction algorithm is implemented in Python, with Keras and Tensorflow as backend.

The values of hyperparameters are used to control the learning process. The LSTM model is tested with various hyperparameters in this case. In order to train and predict the target with greater accuracy, it is critical to select the appropriate hyperparameters combinations. LSTM architectures are chosen and tested, as is the model with the best hyperparameters.

The LSTM hidden layers are limited to 3 and the number of epochs and neurons is set to 200 and 100 respectively. These are some of the hyperparameters that must be chosen and can affect the tradeoff between prediction accuracy and training time. A greater number of layers may improve prediction precision. Based on the relationship between the number of previously observed values and the accuracy of the prediction, which determines the amount of information that the network needs to memorize and use. We use the first 29 weeks of data for training and 7 weeks for testing/validation. The “Rectified Linear Unit (ReLU)” and “ADAM” are used as the activation function and the optimization algorithm respectively. Table 4.3 shows the list of hyperparameters and their values.

Hyperparameters	Values
Hidden Layer	3
Number of Neurons	100
Number of Epoch	200
Learning rate	0.001
Bach size	64
Dropout	0.2
Optimizer	Adam
Activation function	Rectified Linear Unit (ReLU)

Table 4. 3 Hyperparameters and values for LSTM model

4.5 Performance Metrics

Previous papers select several indicators to measure predictive accuracy performance. In this study, the three classical performance metrics (i.e., RMSE, MAE, and MAPE) are selected to measure the forecasting accuracy of each model. The definitions of these performance metrics are as follows:

1. Root Mean Square Error

The Root-Mean-Square Error (RMSE) is a popular metric for assessing the accuracy of a model's prediction. It calculates the differences or residuals between the actual and predicted values. The main advantage of RMSE is that it penalizes large errors. The formula is described in Equation (4.1).

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(xi - xj)2} \qquad (4.1)$$

Where,

- N is the total number of sample observation;
- xi is the actual value;
- xj is the predicted value.

2. Mean Absolute Error(MAE)

The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. The MAE is the average over the verification sample of the absolute values of the differences between forecast and the corresponding observation. The MAE is a linear score which means that all the individual differences are weighted equally in the average. The formula is given in Equation (4.2).

$$\mathrm{MAE} = \frac{1}{N}\sum_{j=1}^{N}|\hat{x}j - Xj| \qquad (4.2)$$

Where;

- N is the total number of sample observation;
- xj is the actual value;
- $\hat{x}j$ is the predicted value.

3. Mean Absolute Percentage Error

The Mean Absolute Percentage Error (MAPE) is a measure of a forecast system's accuracy. It calculates the average absolute percent error for each time period minus actual values divided by actual values to calculate this accuracy as a percentage. The formula for MAPE is given in Equation (4.3).

$$\text{MAPE} = \frac{100\%}{N}\sum_{j=1}^{N}\left|\frac{\bar{x}_j - X_j}{x_j}\right| \qquad (4.3)$$

Where;

- N is the total number of sample observation;
- xj is the actual value;
- $\bar{x}_j$ is the predicated value.

4.6 Results and Discussion

4.6.1 Result Analysis

This study aims to accurately forecast the incoming call volumes traffic based on past observation and make the prediction to be used as an input for operations resource planning. Hence determining resource requirements and scheduling of staff at a call center require accurate forecasts. The result of our analyzer based on the total incoming call volumes traffic as input parameters and forecast future incoming calls.

From the total data set, 273 days of incoming call volumes traffic, first 218 days or 29 weeks' data are used for the training purpose. The remaining 55 days, we use for validation of the model for all models. Figure 4.9 shows the incoming call volume traffic is forecasted for one week ahead with SARIMA (1, 1, 1) (1, 0, 1) 32 model. Based on Figure 4.10, the LSTM model predicts the future incoming call volumes traffic better than the SARIMA model. This better prediction is because the LSTM model can well learn data structure and pattern, such as long term dependence and complexity (nonlinearity)in time series forecasting [34][35].

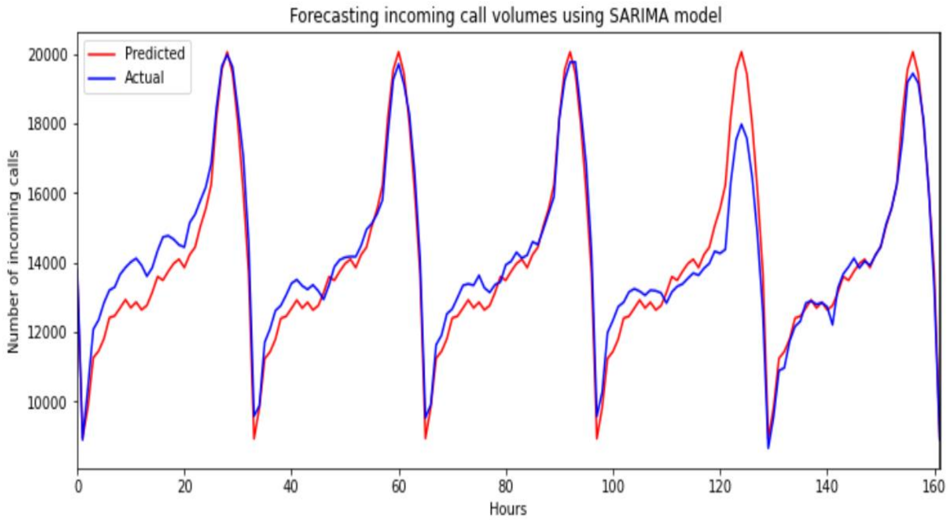


Figure 4. 8 SARIMA Model forecasting for incoming call volumes

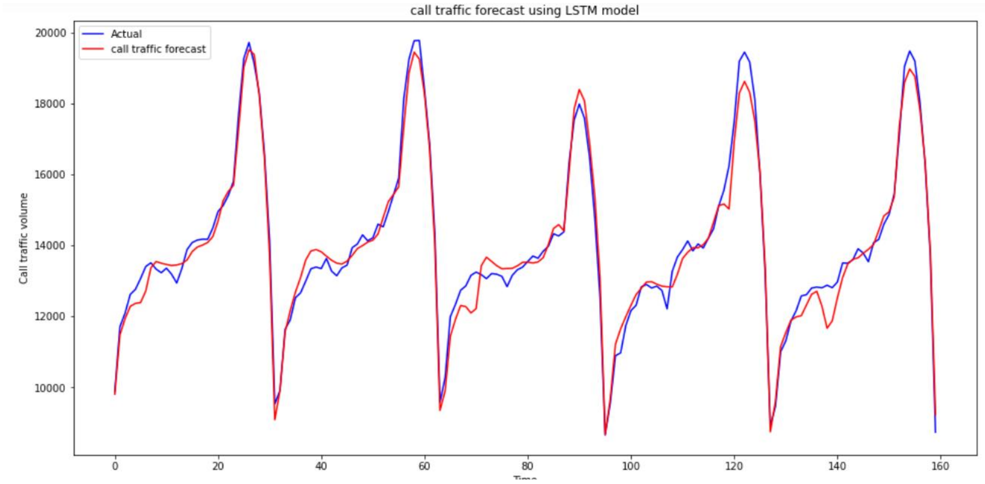


Figure 4. 9 LSTM Model forecasting incoming call volumes

The forecasting accuracy evaluation of the two models is shown in Figure 4.11. The Figure shows that deep learning, LSTM, forecasting models have the lowest forecasting error when compared to the SARIMA model.

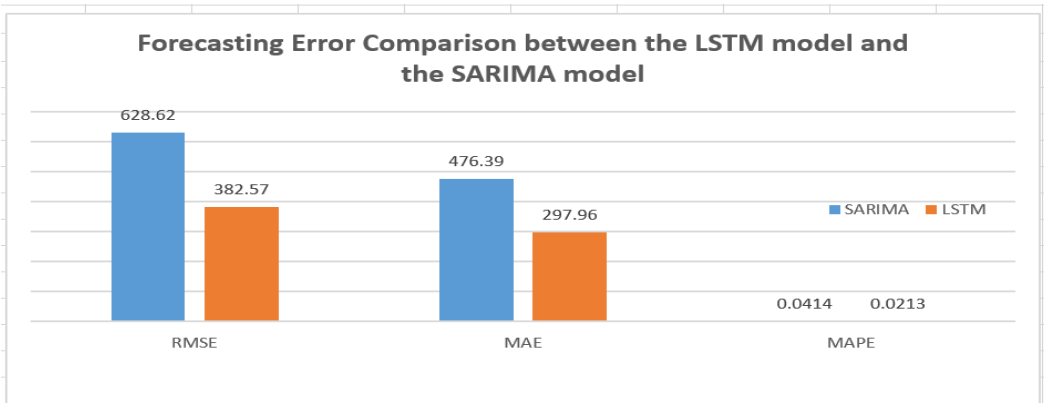


Figure 4. 10 Forecasting error comparison between the LSTM and the SARIMA models

4.6.2 Discussion

There are different statistical forecasting time series models that predict future value based on past observation. However, this model works only the linear part of the traffic and homogeneous time series. The machine learning model incorporates the nonlinear elements of the incoming call volumes traffic but did not work for long term dependence and nonlinear characteristics. Among deep learning models, the LSTM model is used for the prediction of incoming call volume traffic because it handles the character of the incoming call volume traffic such as long term dependency and nonlinear characteristics.

In this study, as indicated in Figure 4.10, the LSTM model presents better performance than the SARIMA model in all error measures. The predicted RMSE of the SARIMA model is 628.6283, while the predicted RMSE of the LSTM model is 382.5707, and the prediction result is enhanced by 24.6% compared to the SARIMA model. When we compare the prediction error using MAE, the SARIMA model's MAE is 476.3979, while the predicted MAE of the LSTM model is 297.9663, so the prediction result is enhanced by 17.8% compared to the SARIMA model. Lastly, the predicted MAPE of the SARIMA, which expresses accuracy as a percentage of error, is 4.14%. In comparison, the MAPE of the LSTM model is 2.13%, so the prediction result is improved by 2.01% compared to the SARIMA model. In general, in all performance metrics, the LSTM model has minimum values of error.

CHAPTER FIVE

Conclusion and Recommendation

5.1 Conclusion

Nowadays, call center services are becoming very critical for many middle and large scale business companies, especially in telecommunication industry. In call center, efficiently managing of incoming call traffic is a big headache due to its dynamic behavior of the customers' request. That means it is not easy to forecast the incoming call traffic in order to assign appropriate resources including staff scheduling. Thus, an algorithm that better understands the complex and non-linearity relationships of incoming call volumes traffic is required to improve forecast accuracy.

The aim of this study is to build a model that forecasts incoming call center call volume traffic using historical data in Ethio Telecom. We have used nine months with a total data record of 8800 incoming call traffic to build the model. It is split into 80% training set and 20% testing set. We have used two univariate time series techniques, namely SARIMA and LSTM to build the model and evaluated to forecast incoming call center call volume traffic.

The experimental results show that the LSTM model has lower RMSE, MAE, and MAPE error metrics when compared with the SARIMA model. An average of 24.6% of RMSE improvement is found. The LSTM model includes more features in incoming call volume traffic that is not incorporated by the SARIMA model. Thus, the overall result demonstrates that LSTM approach could be considered as an effective method for forecasting incoming call volume traffic to reflect temporal patterns and enhance the allocation of resources to decrease the number of incoming call drops.

5.2 Recommendation

We recommend Ethio Telecom to use this model in order to do the appropriate staff level and scheduling for optimal resource utilization and increase customer satisfaction. It can also help future researchers to:

- Build a model using large amount data to increase the accuracy of the model.
- Further study on multivariate time series incoming call traffic to enhance the forecasting on specific call types.

Reference

[1] A. Mandelbaum, "CALL CENTERS (CENTRES) Research Bibliography with Abstracts," 2004.

[2] R. Ibrahim, H. Ye, P. L. Ecuyer, and H. Shen, "Modeling and forecasting call center arrivals : A literature survey and a case study," *Int. J. Forecast.*, no. 2009, 2015, doi: 10.1016/j.ijforecast.2015.11.012.

[3] O. M. Society, R. F. Yolu, and S. Francisco, "The Modern Call Center : A Multi-," vol. 16, no. 6, pp. 665–688, 2008, doi: 10.3401/poms.

[4] N. Mehrbod, A. Grilo, and A. Zutshi, "Caller-Agent Pairing in Call Centers Using Machine Learning Techniques with Imbalanced Data," 2018.

[5] D. Chicu and G. Ryan, "Determinants of customer satisfaction in call centres," vol. 2, no. 2, pp. 20–41, 2016.

[6] B. Models, R. Ibrahim, and P. L. Ecuyer, "Forecasting Call Center Arrivals :A COMPARATIVE STUDY," no. 2010, pp. 1–33, 2011.

[7] E. R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, "USING SIMULATION TO EVALUATE CALL FORECASTING ALGORITHMS FOR INBOUND CALL CENTER Guilherme," pp. 1132–1139, 2013.

[8] "Modeling Daily Arrivals to a Telephone," vol. 50, no. 7, pp. 896–908, 2004, doi: 10.1287/mnsc.1040.0236.

[9] H. Shen and J. Z. Huang, "Analysis of call centre arrival data using singular value decomposition," no. August 2004, pp. 251–263, 2005, doi: 10.1002/asmb.598.

[10] H. Shen and J. Z. Huang, "Forecasting Arrivals to a Telephone Call Center," no. January 2003, 2014.

[11] S. PAI, "Ethio Telecom reports a 22% jump in subscribers to 56.2m." https://developingtelecoms.com/telecom-technology/consumer-ecosystems/11518-ethio-telecom-reports-a-22-jump-in-subscribers-to-56-2m.htm.

[12] T. Engineering, "FORECASTING DEMAND FOR A TELEPHONE CALL CENTER," vol. 10, no. 3, pp. 266–277, 2020.

[13] B. Sennaroglu and G. Polat, "Time series forecasting for a call centre," *Proc. Int. Conf. Ind. Eng. Oper. Manag.*, vol. 2017, no. JUL, pp. 464–468, 2017.

[14] J. W. Taylor and J. W. Taylor, "Forecasting Intraday Arrivals at a Call Center A Comparison of Univariate Time Series Methods for Forecasting Intraday Arrivals at a Call Center," no. September 2015, 2008, doi: 10.1287/mnsc.1070.0786.

[15] T. Chanpanit, "Big Data Framework for Incoming Calls Forecasting in a Call Center," no. June, pp. 12–13, 2020.

[16] D. K. Barrow, "Forecasting intraday call arrivals using the seasonal moving average method," *J. Bus. Res.*, vol. 69, no. 12, pp. 6088–6096, 2016, doi: 10.1016/j.jbusres.2016.06.016.

[17] E. Chromy, M. Kavacky, and I. Baronak, "Call Center optimization - step by step," pp. 119–123, 2017.

[18] S. Kurniali, "Customer Service Information System for a Call Center," *Procedia - Procedia Comput. Sci.*, vol. 59, no. Iccsci, pp. 298–304, 2015, doi: 10.1016/j.procs.2015.07.569.

[19] M. Christian, D. Bdd, and N. Lenaugasse, "Call Centre Architecture and Solutions," 1999.

[20] "An Introductory Study on Time Series Modeling and Forecasting Ratnadip Adhikari R. K. Agrawal."

[21] H. Zou and Y. Yang, "Combining time series models for forecasting," vol. 20, pp. 69–84, 2004, doi: 10.1016/S0169-2070(03)00004-9.

[22] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting , Second Edition Springer Texts in Statistics*. .

[23] I. Khandelwal, R. Adhikari, and G. Verma, "Time Series Forecasting using Hybrid ARIMA and ANN Models based on DWT Decomposition," *Procedia - Procedia Comput. Sci.*, vol. 48, no. Iccc, pp. 173–179, 2015, doi: 10.1016/j.procs.2015.04.167.

[24] B. De Constance and B. De Constance, "Jenkins methodology."

[25] M. Milenković, L. Švadlenka, V. Melichar, and N. Bojović, "SARIMA modelling approach for railway passenger flow forecasting," vol. 4142, no. April, 2016, doi: 10.3846/16484142.2016.1139623.

[26] P. P. Dabral and M. Z. Murry, "Modelling and Forecasting of Rainfall Time Series Using SARIMA," 2017, doi: 10.1007/s40710-017-0226-y.

[27] U. Yolcu, E. Egrioglu, and C. H. Aladag, "A new linear & nonlinear arti fi cial neural network model for time series forecasting," *Decis. Support Syst.*, vol. 54, no. 3, pp. 1340–1347, 2013, doi: 10.1016/j.dss.2012.12.006.

[28] I. Khandelwal, R. Adhikari, and G. Verma, "Time series forecasting using hybrid arima and ann models based on DWT Decomposition," in *Procedia Computer Science*, 2015, vol. 48, no. C, pp. 173–179, doi: 10.1016/j.procs.2015.04.167.

[29] J. Bayer and C. Osendorfer, "Learning Stochastic Recurrent Networks," pp. 1–9, 2014, [Online]. Available: http://arxiv.org/abs/1411.7610.

[30] H. Palangi, R. Ward, and L. Deng, "Distributed Compressive Sensing: A Deep Learning

Approach," *IEEE Trans. Signal Process.*, vol. 64, no. 17, pp. 4504–4518, 2016, doi: 10.1109/TSP.2016.2557301.

[31] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLoS One*, vol. 12, no. 7, pp. 1–24, 2017, doi: 10.1371/journal.pone.0180944.

[32] E. Chromy, T. Misuth, and M. Kavacky, "E c f i u i c c," vol. 9, no. 1, 2011.

[33] T. Models, "Introduction to Traffic Modeling and Resource Allocation in Call Centers," 1917.

[34] S. Siami-namini and N. Tavakoli, "A Comparison of ARIMA and LSTM in Forecasting Time Series," pp. 1394–1401, 2018, doi: 10.1109/ICMLA.2018.00227.

[35] A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo, "Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction," vol. 2014, pp. 9–11, 2014.

Appendix

Time Series Forecasting for Incoming Call Volume using LSTM: the case of Ethio Telecom Call Center

Mebrhit Hailay

School of Electrical and Computer Engineering

Addis Ababa University

Addis Ababa, Ethiopia
mehailay@gmail.com

Dereje Hailemariam

School of Electrical and Computer Engineering

Addis Ababa University

Addis Ababa, Ethiopia
derejehmr@gmail.com

***Abstract*— An accurate forecasting of incoming call volume traffic is key for the operational planning of an inbound call center service. By forecasting, the correct number of incoming calls volumes traffic, can determine staffing and scheduling levels, improve service requirements, and meet customer satisfaction. The purpose of this study is to build a model that forecasts incoming call center call traffic using Ethio Telecom call center historical data. In this thesis, we proposed time series forecasting model for forecasting the incoming call volume traffic. We have used two univariate time series techniques, namely SARIMA and LSTM. Nine months of incoming call center call traffic data is collected from Ethio Telecom. Finally, experimental result indicates that the LSTM model has 24.6% of RMSE improvement of forecasting error compared to the SARIMA model. The overall results of this research work demonstrate that the LSTM model is an effective method for predicting incoming call volume traffic to reflect temporal patterns.**

***Keywords*— call center, time series, forecasting, Staffing, scheduling, statistical model, deep neural network, LSTM**

Introduction

Many businesses are establishing their own call center service departments these days. With technological advancements, managing call center service efficiently and effectively is critical in terms of economic interest, providing quality of service (QoS) and improving customer experience[1] . Call center is a service providing department which is used as a frontline that allows customers to request and get support directly from the company. It is one of the most important forms of communication between businesses and their customers. Call centers can be used in different business areas such as banks, airlines, healthcare, and telecommunications to handle a high volume of customer call requests [2] [3].

The increasing number of businesses encourages companies to have their own call center to provide the best possible services and information to their customers [4][5]. Customers save significant time by using call centers efficiently, while businesses run significantly more efficient operations. In order to improve customer satisfaction, call center managers face numerous challenges in providing high-quality service while keeping operating costs low. Because managers are continually presented with tremendous uncertainty, effective call center management is a difficult undertaking. Because it is often time-varying, stochastic, reliant on time periods and call kinds, and frequently influenced by external events, the number of incoming calls is a substantial source of uncertainty. Accurate modeling and forecasting of future call arrival volumes is a complicated subject that is vital for important call center operational decisions [6].

Mainly there are two types of call centers: inbound and outbound call centers. Inbound call centers are the most common type of call centers, where customers initiate calls and call center agents respond to those calls. Outbound call centers are another type of call center in which the company makes outgoing calls to customers [3].

The data generated from call centers have a significant amount that should be analyzed and used for decision-making purposes in order to determine appropriate staff levels, shift schedules, and real-time routing design future demands [7] [8].

As call center incoming call volume traffic forecasting is critical, several authors have been searching better call center traffic forecasting methods. In [6], different statistical time series methods for forecasting future call volumes in call centers are evaluated and compared. They consider different forecasting models including the Fixed-Effects (FE) Model, the Gaussian Linear Mixed Effects (ME) Model, the Holt-Winters (HW) Smoothing model and the “Top-Down” (TD) Approach. The main motivation of the study is to solve the challenges faced by call center manager such as the problem of determining appropriate staffing levels, the problem of scheduling (and re-scheduling) the available pool of agents based on updated forecasts and routing incoming calls in real-time to available agents in the major company of Canada. In the study [11], the authors have forecasted the number of incoming call center calls using dynamic time series three forecast methods namely, Moving Averages method, Simple Exponential Smoothing method and Additive HW method have been used.

In [12] mainly evaluates univariate time series methods: SARIMA modeling; robust exponential smoothing based on exponentially weighted least absolute deviations regression; dynamic harmonic regression, which is a form of unobserved component state-space modeling for forecasting intraday arrivals, periodic autoregressive modeling; and an extension of HW exponential smoothing for the case of two seasonal cycles, for lead times from half-hour ahead to two weeks ahead. The authors in [13] have proposed a Big Data framework for forecasting incoming calls to the mobile call center. They conducted data analysis using historical data of incoming calls from 2015-2018 years and defined two models and flow diagrams.

The study in [14] determines the optimal number of agents in contact center service based on downtime and administrative task duration parameters. The authors have proposed a mathematical model for contact centers and a six step estimation model to compute the optimal number of contact. Finally, they concluded the Erlang C formula tool is suitable for modeling the QoS parameters of contact centers. The authors in [15] investigated time series forecasting methods to effectively manage the bank's workforce call center, and the seasonal moving average (SMA) method is used for systematic evaluations, as are artificial neural networks (ANNs) for forecasting intraday call arrivals.

All of these researches are based on statistical analysis models that better handle time series data having linear relationships. However, incoming calls of call centers have complex and non-linear behaviors, and their accurate forecast is very critical to efficiently performs staffing and scheduling. So, an algorithm that better understands the complex and non-linearity relationships of inbound customer calls is required to improve forecast accuracy.

The paper is organized as follows, section II discusses methodology (time series call volume modeling and forecasting techniques). The experimental setup is discussed in section III. Section IV describes the results and discussion of an experiment. Finally, Section V discusses the conclusion and recommendation.

Methodology

In this study, we proposed two univariate time series forecasting methods: statistical and deep learning methods for forecasting the incoming call volume traffic. A time series forecasting method is a technique for predicting the future value based on current and historical data.

SARIMA Model

SARIMA model is a statistical forecasting model that uses time series data to either better understand the data set or to predict future trends. SARIMA models are an extension of ARIMA that explicitly supports univariate time series data with seasonal components. This model is divided into two parts: non-seasonal and seasonal, with the Autoregressive (AR), differencing (d), and moving average (MA) parameters in each. It is used when there is a periodic characteristic in the data that must be known ahead of time. SARIMA (p,d,q)(P,D,Q)s is the general form of a seasonal ARIMA model, where p is the non-seasonal AR order, d is the non-seasonal differencing, q is the non-seasonal MA order, P is the seasonal AR order, D is the seasonal differencing, Q is the seasonal MA order, and s is the time steps of repeating seasonal pattern.

The most crucial step in estimating the SARIMA model is determining the values of seven parameters (p,d,q)(P,D,Q,s). If the variance grows with time, for example, based on the time plot of the data, we should use variance-stabilizing transformations and differences. Then, using the ACF to determine how much linear dependence exists between observations in a time series separated by a lag p and the PACF to determine how many autoregressive terms q is required, we can identify the preliminary values of autoregressive order p, order of differencing d, moving average order q, and their corresponding seasonal parameters P, D, and Q.

LSTM Model

The LSTM model is a type of recurrent neural network that can remember patterns selectively for long periods of time. The hidden layer unit in the original RNN architecture is replaced by a memory blocks (called cells) in the LSTM architecture. It carries all the information with only some linear interaction. The LSTM does have the ability to remove or add information to the cell state, carefully regulated by structures called gates. The gates, which are based on a sigmoid neural network layer, allow cells to either pass data through or discard it. Each sigmoid layer produces a number between 0 and 1, indicating how much of each data segment should be allowed through in each cell. A value of zero means "let nothing through," while a value of one means "let everything through!" [18]. The LSTM network unit consists of a cell, an input gate, an output gate, and a forget gate. A cell retains values for an autocratic time interval. The input gate controls the flow of information into the cell. The output gate controls the flow of information to and from the outside world. Similarly, forget gates control the flow of information that is necessary or unnecessary.

Experimental Setup

Data Set

The dataset used in this study were collected from the inbound call center operations IPCC system of Ethio telecom. It is collected for over 273 days from August 1, 2020 to April 30, 2021, on a half hourly basis. The call center operates 16 hours per day from 7:00AM to 11:00, PM seven days a week. The total number of recorded observations 8,736 data set.

Data Preprocessing

The collected data is preprocessed. Firstly, data is visualized, and handling missing values and data normalization are performed. In the data analysis, MS Excel and the Python tools are used. There are 3 hours missing values in the data, one hour missing value on August 20, 2020, and the other two hour January 28, 2021. The missing values are filled using a linear interpolation method. The dataset is divided into two groups: training set (80%) to training SARIMA and LSTM algorithms and testing set (20%) for evaluate the performance of the forecasting accuracy of SARIMA and LSTM algorithms.

Performance Metrics

Three performance metrics, such as Root-Mean-Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error

(MAPE), are used to compare the performance of the different models developed in the thesis for measuring the accuracy of the prediction algorithm. Forecasting models with the minimum of these performance metrics are considered the best.

Algorithms for Call Volume Forcasting

1. The SARIMA Model

SARIMA models are an extension of ARIMA that explicitly supports univariate time series data with a seasonal component. The most crucial step in estimating the SARIMA model is determining the values of seven parameters (p,d,q)(P,D,Q,s). To determine the value of SARIMA model parameters using ACF and PACF plots and grid search. TABLE I shows model summary of the SARIMA parameter values.

2. The LSTM Model

The LSTM is a RNN that can store and learn from long sequences of observations. The values of hyperparameters are used to control the learning process. The LSTM model is tested with various hyperparameters in this case. LSTM architectures are chosen and tested, as is the model with the best hyperparameters. TABLE II shows the model summary of LSTM parameter values.

TABLE I. MODEL SUMMARY OF SARIMA MODEL HYPERPARAMETERS AND VALUES

Hyperparameters	Values	Hyperparameters	Values
p	1	P	1
d	1	D	0
q	1	Q	1
		s	32

TABLE II. MODEL SUMMARY OF LSTM HYPERPARAMETERS AND VALUES

Hyperparameters	Values	Hyperparameters	Values
Hidden Layer	3	Bach size	64
Number of Neurons	100	Dropout	0.2
Number of Epoch	200	Optimizer	Adam
Learning rate	0.001	Activation function	ReLU

Results and Discussion

This study aims to accurately forecast the incoming call volumes traffic based on past observation and make the prediction to be used as an input for operations resource planning. Hence determining resource requirements and scheduling of staff at a call center require accurate forecasts. From the total data set, 273 days of incoming call volumes traffic, first 218 days or 29 weeks' data are used for the training purpose. The remaining 55 days, we use for validation of the model for all models. Fig 1 and Fig 2 shows the incoming call volume traffic forecasted for one week ahead using SARIMA and LSTM model respectively.

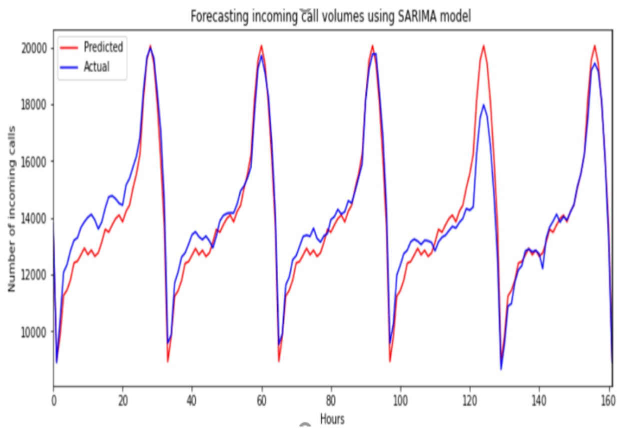


Fig. 1. SARIMA model forecasting for incoming call volume

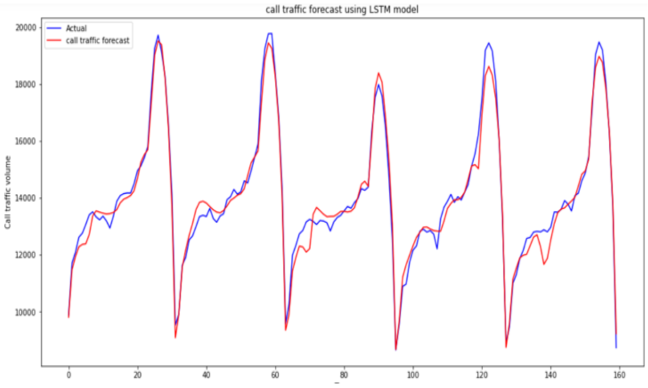


Fig. 2. LSTM model forecasting for incoming call volume

The accuracy of the SARIMA and LSTM forecasting methods can be compared by RMSE, MAE and MAPE performance metrics. The summary of performance metric values of these tests are shown in TABLE III.

TABLE III. SUMMARY OF PERFORMANCE METRICS VALUES WITH SARIMA AND LSTM MODELS

Model	RMSE	MAE	MAPE
SARIMA	628.62	476.39	4.14%
LSTM	382.57	297.96	2.13%

In this study, as indicated in Fig 2, the LSTM model presents better performance than the SARIMA model in all error measures. The predicted RMSE of the SARIMA model is 628.6283, while the predicted RMSE of the LSTM model is 382.5707, and the prediction result is enhanced by 24.6% compared to the SARIMA model. When we compare the prediction error using MAE, the SARIMA model's MAE is 476.3979, while the predicted MAE of the LSTM model is 297.9663, so the prediction result is enhanced by 17.8% compared to the SARIMA model. Lastly, the predicted MAPE of the SARIMA, which expresses accuracy as a percentage of error, is 4.14%. In comparison, the MAPE of the LSTM model is 2.13%, so the prediction result is improved by 2.01% compared to the SARIMA model. In general, in all performance metrics, the LSTM model has minimum values of error.

Conclusion and Recommendation

Nowadays, call center services are becoming very critical for many middle and large scale business companies, especially in telecommunication industry. In call center, efficiently managing of incoming call traffic is a big headache due to its dynamic behavior of the customers' request. That means it is not easy to forecast the incoming call traffic in order to assign appropriate resources including staff scheduling. Thus, an algorithm that better understands the complex and non-linearity relationships of incoming call volumes traffic is required to improve forecast accuracy. The aim of this study is to build a model that forecasts incoming call center call volume traffic using historical data in Ethio Telecom. We have used two univariate time series techniques, namely SARIMA and LSTM to build the model and evaluated to forecast incoming call center call volume traffic. The experimental results show that the LSTM model has lower RMSE, MAE, and MAPE error metrics when compared with the SARIMA model. An average of 24.6% of RMSE improvement is found.

We recommend Ethio Telecom to use LSTM model in order to do the appropriate staff level and scheduling for optimal resource utilization and increase customer satisfaction. It can also help future researchers to:

- Build a model using large amount data to increase the accuracy of the model.
- Further study on multivariate time series incoming call traffic to enhance the forecasting on specific call types.

References

[1] A. Mandelbaum, "CALL CENTERS (CENTRES) Research Bibliography with Abstracts," 2004.

[2] R. Ibrahim, H. Ye, P. L. Ecuyer, and H. Shen, "Modeling and forecasting call

center arrivals : A literature survey and a case study," Int. J. Forecast., no. 2009, 2015, doi: 10.1016/j.ijforecast.2015.11.012.

[3] O. M. Society, R. F. Yolu, and S. Francisco, "The Modern Call Center : A Multi-," vol. 16, no. 6, pp. 665–688, 2008, doi: 10.3401/poms.

[4] N. Mehrbod, A. Grilo, and A. Zutshi, "Caller-Agent Pairing in Call Centers Using Machine Learning Techniques with Imbalanced Data," 2018.

[5] D. Chicu and G. Ryan, "Determinants of customer satisfaction in call centres," vol. 2, no. 2, pp. 20–41, 2016.

[6] B. Models, R. Ibrahim, and P. L. Ecuyer, "Forecasting Call Center Arrivals :A COMPARATIVE STUDY," no. 2010, pp. 1–33, 2011.

[7] E. R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, "USING SIMULATION TO EVALUATE CALL FORECASTING ALGORITHMS FOR INBOUND CALL CENTER Guilherme," pp. 1132–1139, 2013.

[8] "Modeling Daily Arrivals to a Telephone," vol. 50, no. 7, pp. 896–908, 2004, doi: 10.1287/mnsc.1040.0236.

[9] H. Shen and J. Z. Huang, "Analysis of call centre arrival data using singular value decomposition," no. August 2004, pp. 251–263, 2005, doi: 10.1002/asmb.598.

[10] H. Shen and J. Z. Huang, "Forecasting Arrivals to a Telephone Call Center," no. January 2003, 2014.

[11] T. Engineering, "FORECASTING DEMAND FOR A TELEPHONE CALL CENTER," vol. 10, no. 3, pp. 266–277, 2020.

[12] J. W. Taylor and J. W. Taylor, "Forecasting Intraday Arrivals at a Call Center A Comparison of Univariate Time Series Methods for Forecasting Intraday Arrivals at a Call Center," no. September 2015, 2008, doi: 10.1287/mnsc.1070.0786.

[13] T. Chanpanit, "Big Data Framework for Incoming Calls Forecasting in a Call Center," no. June, pp. 12–13, 2020.

[14] E. Chromy, M. Kavacky, and I. Baronak, "Call Center optimization - step by

step," pp. 119–123, 2017.

[15] D. K. Barrow, "Forecasting intraday call arrivals using the seasonal moving average method," J. Bus. Res., vol. 69, no. 12, pp. 6088–6096, 2016, doi: 10.1016/j.jbusres.2016.06.016.

[16] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," PLoS One, vol. 12, no. 7, pp. 1–24, 2017, doi: 10.1371/journal.pone.0180944.