# ETHIOPIAN HEALTH COMMODITY DEMAND PREDICTION

**AYELE TAMIRU BIRHANE**

**A Thesis Submitted to the Department of Computer Science and Engineering**

**School of Electrical Engineering and Computing**

**Presented in Partial Fulfillment of the Requirement for the Degree of Masters in Information System Engineering**

**Office of Graduate Studies**

**Adama Science and Technology University**

**Adama, Ethiopia**

**October,2019**

# ETHIOPIAN HEALTH COMMODITY DEMAND PREDICTION

**AYELE TAMIRU BIRHANE**

**Advisor: DR. TILAHUN MELAK SITOTE**

**A Thesis Submitted to the Department of Computer Science and Engineering**

**School of Electrical Engineering and Computing**

**Presented in Partial Fulfillment of the Requirement for the Degree of Masters in Information System Engineering**

**Office of Graduate Studies**

**Adama Science and Technology University**

**Adama, Ethiopia**

**October,2019**

# Approval of Board of Examiners

We, the undersigned, members of the Board of Examiners of the final open defense by **Ayele Tamiru Birhane** have read and evaluated his/her thesis entitled "***Ethiopian health commodity demand prediction"*** and examined the candidate. This is, therefore, to certify that his thesis has been accepted in partial fulfillment of the requirement of the Degree of Masters of Science in Information Systems Engineering.

| | | |
|---|---|---|
| _____ | _____ | _____ |
| Advisor | Signature | Date |
| _____ | _____ | _____ |
| Chair Person | Signature | Date |
| _____ | _____ | _____ |
| Internal Examiner | Signature | Date |
| _____ | _____ | _____ |
| External Examiner | Signature | Date |

# DECLARATION

I hereby declare that this MSc thesis is my original work, has not been presented for a degree in any other university, and all source of materials used for this thesis have been properly acknowledge.

Name: Ayele Tamiru

Signature _____

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

Name: Dr. Tilahun Melak Sitote

Signature_____

Date of Submission: October,2019

# Dedication

I dedicate my thesis work for my father Tamiru Birhane and Sinkenesh Kefeyalew.

# ACKNOWLEDGMENT

Firstly, it is my great endeavor to offer my deepest gratitude to my Almighty lord for the wisdom he bestowed upon me, the strength, peace of my mind, and good health in order to finalize this research. Secondly, my deepest gratitude goes to my advisor, Dr. Tilahun Melak Sitote for his uninterrupted effort to advise me in all aspects.

Thirdly my Special thanks should be given to Eyerus Urge for her love and unlimited consistent kindness. I need to give a very special thanks for those who assist and support me in providing some basic information's and interpretation related for some common professional words regarding on the area of the study, Mr. Tagay Demie, Mr. Sisay. Miss. Mehert and Tewodros Asmare and to all my class mates.

Last, I would like to acknowledge Sinkenesh Kefeyalew and Tamiru Birhane for their pure love as well as brother and sister for unlimited support.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ANN | Artificial Neural Network |
| ARIMA | Autoregressive Integrated Moving Average |
| ARMA | Auto-Regressive Moving Average |
| ARMDN | Associative and Recurrent Mixture Density Networks |
| BFGS | Fletcher-Gold-Shannon |
| BP | Backpropagation |
| BP | Back Propagation |
| BWE | Bullwhip Effect Exists |
| BWE | Bullwhips Effect |
| CFPR | Collaborative Forecasting Planning & Replenishment |
| CSV | Comma Delimited |
| DNN | Deep Neural Networks |
| DT | Decision Tree |
| DWT | Discrete Wavelet Transforms |
| DWT-ANN | Artificial Neural Network |
| EPSA | Ethiopian Pharmaceuticals Supply Agency |
| EPSA | Ethiopian Pharmaceuticals and Supply Agency |
| FMOH | Federal Ministry of Health |
| GA | Genetic Algorithms |
| GD | Gradient Descent |
| GTB | Gradient Tree Boosting |
| GTP | Growth and Transformation Plan |
| HCMIS | Health Commodity Management Information System |
| IPLS | Integrated Pharmaceuticals Logistics System |
| LM | Learning Machine |
| LMIS | Logistics Management Information System |
| LSI | Look Seasonality Indica |
| LSTM | Long Short-Term memory |
| MAPE | Mean Average Percentage Error |
| MAPE | Mean Absolute Percentage Error |
| MLP | Multi-Layer perceptron |

| | |
|---|---|
| MSE | Mean Square Error |
| NNs | Neural Networks |
| NNs | neural networks |
| NPPL | National Pharmaceuticals Procurement List |
| NMSE | Normalized Mean Square Error |
| NSAmp | Net Stock Amplification |
| NWMSLE | Normalized Weighted Mean Square Log Error |
| PE | Processing Elements |
| PLMP | Pharmaceuticals Logistics Master Plan |
| R | R program |
| RBF | Radial Basis Function |
| RDF | Revolving Distribution Fund |
| MSE | Mean Square Error |
| RNN | Recurrent Neural Network |
| RNN | Recurrent Neural Network |
| SOMs | Self-Organizing Maps |
| SOP | standard operating procedures |
| STV | Stock Transfer Voucher |
| SVM | Clustering, Support Vector Machines |
| SVMs | Support Vector Machines |
| UA | Updated Accumulators |
| USD | United States Dollar |
| VMI | Vendor Managed Inventory |

# ABSTRACT

*A pharmaceutical agency's ability to forecast demand accurately is requisite to identify, rationalize and improve results. The optimization of order quantity, stock level, or delivery schedule depends on forecasting accurately, demand at the store level. Forecasting is important in the perspective of the pharmaceutical areas, which commonly requires efficient Supply Chain Management. The main objective of study was to forecast demand by using EPSA consumption data. EPSA consumption data was extracted from HCMIS database from the year 2016 to 2019. After the data is converted into model understandable format, comparative study is done on three algorithms (LSTM, Multi-Layer Perceptron and Auto ARIMA). The best performing algorithm is used to forecast the future. By using MSE metric, we conclude that Auto ARIMA gives the best prediction performance from the three algorithms. It also has a closer result comparing with the actual data. The MSE results for Auto ARIMA in the selected four items are Atropine Sulphate - 1mgml - Injection (Sulphate) is 329, Tetanus Antitoxin (Human) Equine - 1500 Units – Injection is 3,747, Insulin Isophane Human - 100IUml in 10ml Vial – Injection (Suspension) is 36,290 and Dextrose - 40% in 20ml - Intravenous Infusion is* 173,979 *respectively. Predicting demands accurately also helps to accommodate time, cost, flexibility, space, efficient inventory management, wastage reduction and minimize product expiry properly.*

***Keywords:*** *Demand forecasting, Consumption, Time Series Forecasting.*

# CHAPTER ONE: INTRODUCTION

## 1.1. Background of the Study

Underestimating demand can result in lost sales, dissatisfied customers and insufficient resources. Overestimating demand can result in high operating costs and a high inventory. Planning decisions and inventory management could be steered by effective forecasting. Effective forecasting is essential to achieve service levels, to plan allocation of total inventory investment, to identify needs for additional production capacity, and to choose between alternative operating strategies [1].

Forecasting is an integrated exercise in which all levels of the supply chain are involved and are willing to share information which helps in increasing demand visibility within organizations as well increase the performance of forecast. Firms that use forecasting successfully had developed not only cross-functional trust, but also cross-organizational trust with distributors and suppliers. Forecasts is critical joints connecting upstream and downstream supply chain processes and activities to deliver the ultimate results. Forecast and its degree of accuracy are making various impacts on supply chain performance. The overall objective of a supply chain is to transform and transport materials or products, add value, and satisfy the demand of each step of the process. Understanding downstream demand is prerequisite to accomplishing the overall supply chain objective [2].

Demand and sales forecasting are one of the most important functions of manufacturers, distributors, and trading firms. Keeping demand and supply in balance, they reduce excess and shortage of inventories and improve profitability. When the producer aims to fulfill the overestimated demand, excess production results in extra stock keeping which ties up excess inventory. On the other hand, underestimated demand causes unfulfilled orders, lost sales foregone opportunities and reduces service levels. Both scenarios lead to inefficient supply chain [3].

Demand forecasts are essential for managing supply chain activities but are difficult to create when collaborative information is absent. Many traditional and advanced forecasting tools are available, but applying them to a large number of

customers is not manageable. Traditional statistical method is less effective when there are many downstream supply chain members exhibiting a variety of usage patterns, particularly when the usage patterns are non-linear [4].

Demand forecasting in pharmaceutical industry has more complex structure than other sectors. Human factors, seasonal and epidemic diseases, market shares of the competitive products and marketing conditions are considered as main external factors for forecasting pharmaceutical product. Additionally, active ingredients rate is also important factor for forecasting process [5].

The traditional supply chain behaviors where customers placed orders and generally advised their demand requirements are no longer prevalent; nowadays, responsibility for inventory management and forecasting has often shifted upstream to the vendor. Collaborative Forecasting Planning & Replenishment and Vendor Managed Inventory are two strategies where forecasting responsibility is transferred to the supplier. Demand forecasting accuracy can be affected by different situation including inefficient data collection and poor system infrastructure [6].

The two major types of forecasting methods are, qualitative and quantitative. Quantitative forecasting method depend on the use of statistical techniques for making projections about the future based on the past, while qualitative analysis is based on critical and expert opinion. Qualitative approaches model demand from solicited opinions. For products (for use in our current study) with consumption history available, future activity can be predicted better using quantitative models from sales in the previous cycle. There are two main types of quantitative forecasting time series and causal. Time series analysis predicts future attributes from the historical past and prior experience. This method uses time as the independent variable to predict demand. The causal relationship is applicable under the assumption that there exists a cause and effect relationship between an input variable and its corresponding output [7].

Currently in Ethiopia the Federal Ministry of Health (FMOH) has been working to guarantee an efficient and high-performing healthcare supply chain that was confirm equitable access to affordable medicines for all Ethiopians. Although

numerous challenges remain an inadequate supply of quality and affordable essential pharmaceuticals, poor storage conditions, and weak stock management resulted in high levels of waste and stock outs. To address these challenges, the FMOH started a comprehensive supply chain strategic planning process, which led to the Ethiopian pharmaceutical supply agency (EPSA), EPSA was established in 2007 [8] by Proclamation No. 553/2007 based on the Pharmaceuticals Logistics Master Plan (PLMP). The Agency is authorized to avail affordable and quality pharmaceuticals sustainably to all public health facilities and ensure their rational use [9]. EPSA supplies essential pharmaceuticals to public health facilities by procuring from local and international markets and purchasing drugs locally manufactured and imported from abroad and stores them in his 19 distribution branches and one central distribution center, also sales pharmaceuticals to different private Hospitals and Distribute health program drugs freely to health institutions such as public hospitals, and facilities based on their request [9].

Poor forecasting is not exclusively an internal issue. Suppliers and customers are also directly affected by forecasting practices. Relationships with suppliers can be significantly deteriorated by poor forecasting. Not only that, but that deterioration can have a further impact on a company's supply chain down the line [10]

EPSA uses HCMIS database to control all the transaction of the health commodity. EPSA collects large dataset by using HCMIS database. Even though, agency don't use the generated data for demand forecasting. The demand forecasting in the agency done using a manual way starting from quantification at facility manual bin card. quantification is the process used to control how much a product is required for the purpose of procurement. In addition to approximating the quantities needed of medicines, quantification should estimate the financial requirements to purchase the medicines. Quantification must include contextual factors, such as available funds, storage capacity, capacity to deliver services, and human resources. The next step after quantification is procurement which search for to ensure the availability of the right medicines in the right quantities, at reasonable prices, and at recognized standards of quality. During the quantification there is high possibility the bin card may not be update and this leads the

quantification data error. The quantification data error affects the forecasting accuracy of the agency. Demand forecasting is a great challenge to date in the agency.

The main objective of this study is to select demand forecasting method by comparing three algorithms Auto ARIMA, Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM).

key strengths, weaknesses, opportunities and threats of crime prevention in Ethiopia.

| **Strengths** | **Weakness** |
|---|---|
| ❖ The ability to use the efficient and effective forecasting method.<br>❖ The increased forecasting accuracy of demand.<br>❖ Highly efficient in supply chain management system<br>❖ The ability to take advantage on market.<br>❖ Increasing skills in introducing innovative and modern solutions on demand forecasting to EPSA. | ❖ Inadequate follow-up on implementation data quality<br>❖ Weak data collection and analysis system<br>❖ Lack of ICT infrastructure at all facility level.<br>❖ Lack of data on demographic data, increasing and decreasing demand event data.<br>❖ Shortage skilled professionals on data analysis |
| **Opportunities** | **Threats** |
| ➢ Developments pharmaceuticals demand forecasting strategy will improve efficiency, ability to avoid stockout and wastage rate.<br>➢ High level of competitiveness compared to other companies with similar profile of activity in the suppliers of pharmaceuticals.<br>➢ Flexible adaptability to existing customer demand in terms of both quantity and quality.<br>➢ The possibility to significantly expand the service and availability of the product<br>➢ Real opportunities to avoid stockout. | ❖ Lack of giving attention for data quality<br>❖ Lack of understanding poor data quality consequence on pharmaceutical demand forecasting<br>❖ Lack of giving attention for expanding ICT infrastructure at facility level. |

## 1.2. Motivation

In EPSA, accuracy of demand forecasting is the main causes of problems associated with in efficient health commodity availability and consumer satisfaction. During demand forecasting agency uses paper based manual method that had a lot of errors during data collection from bin card, aggregation, quantification, and complexity of tasks, huge money loss. Thus, my motivation is to contribute my part in applying demand forecasting method for EPSA and it is my dream to look those problems associated with demand forecasting is resolved.

## 1.3. Statement of the Problem

Misleading quantification of the demand for the chosen solution usually implies that the need for the proposed project is exaggerated. Underestimation of the demand in situations where growth is not considered desirable. Misleading analyses of what will happen if a proposed investment project is not implemented have in some cases left the impression that the proposed solution is necessary in order not to end in a future situation few would wish [11]. To identify our problem, experts were interviewed. During the interviews with these experts, who all work for the supply chain or the distribution department, we discovered cohesion between the problems that were mentioned. The supply chain department is charged with regard to delivery performance to customers, in balance with working capital. To create optimal performance from working capital, the forecast accuracy needs to be improved. Therefore, we identified the following core problem. The supply chain core components including demand forecasting, in developing countries due to poor forecasting accuracy the supply chain activities and process are significantly interrupted and affected, this intern the overall performance create burden, and loss of huge cost. The current demand forecasting held by EPSA is one of the traditional and old method of demand forecasting, it starts using an excel spreadsheet to collect data, analyze and predict the data manually using paper-based quantification.

Currently EPSA has faced great problems in terms of an accurate prediction due un existence of demand forecasting method, and this also one of the main causes

of shortage of drugs, high cost import, expiry of drugs, an necessary drug purchases, an necessary cost is allocating for shipping, loading unloading, transportation cost from boarder to central warehouse, during disposing of expired drug, time consuming during transportation, stock out and over stock, frequent emergency purchase request, and others.

In order to alleviate the above stated problems, it is essential and option to deploy forecasting methods such as machine learning and statistical method that have best demand forecasting performance.

## 1.4. Research Questions

This study fills the gap and can support the EPSA by exploring and selecting appropriate algorithms for forecasting future demand.

2. What methods and ~~or~~ techniques had been used at EPSA to forecast the future demand?
3. What are the major limitations in the current EPSA to give an accurate forecasting for future demand?
4. Which ML algorithms more appropriate to forecast the future demand of EPSA?

## 1.4. Objectives of Research

### 1.4.1. General Objectives

The General objective of the study is to predict the demand of Ethiopian Health Commodity**.**

### 1.4.2. Specific Objectives

To accomplish the aforementioned general objective, the following specific objectives developed:

- To identify the nature of forecast and the demand forecasting process.
- To assess the potential of machine learning and statistical methods in demand forecasting.

- To explore and choose among the various software that support and execute machine learning and statistical techniques with consumption datasets.
- To build and train as well as test the performance of the model.
- To come up with a model that is capable to extract demand forecasting.
- To interpret and analyze the results of the model that how strong is the model to forecast the future demand.
- To compare the machine learning with statistical method and select the one which performs the best.
- Forward recommendation for the further study.

## 1.5. Scope and Limitation

### 1.5.1. Scope

The focuses of this thesis on vital of vital products, all other products are excluded from analysis. Not all vital of vital will be analyzed. During implementation execution, a 4-tracer product will be chosen in cooperation with EPSA experts. This product is potentially lifesaving, the best tool to curl the morbidity of the area: in the absence of this pharmaceutical, the patient may die, or may be hurt. crucial to provide the basic health services; without which it is impossible to deliver the basic services in a specific area. It is mandatory 24 hours of a day, 7 days of a week, and or 12 months of a year. This product easily manageable by the health care team and is the most fit with the clinical setup of the facility than the other alternatives [12]. Therefore, the scope of the study is only focused on selecting best demand forecasting based on Mean Square Error (MSE) method and testing the algorithms using the data that was pre-processed and gathered from EPSA HCMIS database. Finally, the best suitable forecast method assessed by ease of analysis and comparative performance metrics was identified for different categories of Items.

### 1.5.2. Limitation

The study limitation was since the over-all health pharmaceuticals in types are too many, the research only used the most vital of vital pharmaceuticals consumption data. Furthermore, Forecasts for non-vital of vital items are excluded. Also, not

all vital of vital items are included the study focuses on four items and the agency uses those items as tracer.

## 1.6. Research Methods

### 1.6.1. Data Collection

The data comes from EPSA starting from 2016-2019 that is extracted from HCMIS data base. The agency manages items about 1400 items 57 of them are vital of vital items. Therefore, the study selected purposely from out of 57 about 4 items. The data includes about 30,038 records and 13 features. The consumption data was used to this study to achieve the goal of the study.

### 1.6.2. Data Preparation

The data was raw and the study prepared this data to make suitable for preprocessing. The data format was changed from XLSX to CSV format. Features also selected.

### 1.6.3. Data Preprocessing

Data preprocessing is the basic stage in order to preprocess and prepared the data in the way it will be suitable to feed for the algorithms. In this stage data cleaning, missing value handling, data splitting has been performed.

### 1.6.4. Algorithm Selection and Forecasting

While reviewing different researchers had made comparative analysis on different prediction algorithms. ARIMA, LSTM, Multi-Layer Perceptron are best performer in different three papers. So, the researcher selected ARIMA, LSTM and Multi-Layer Perceptron algorithms for comparison. The best algorithm is selected based on Mean Square Error (MSE) value, and the best performed algorithm was selected for forecasting. The forecasting is conducted using the best performer prediction algorithm.

### 1.6.5. Tools

The researcher is used anaconda to conduct this study. Because anaconda contains phyton 3.6, RStudio and Jupyter notebook all in one. Python and Jupyter notebook

used to execute LSTM and Multi-Layer Perceptron algorithms and for visualization, RStudio used to execute ARIMA model and Imputation techniques.

## 1.7. Significance of Study

The purpose of this research was aims to develop predictive model and pattern analytics,

EPSA and different stake holders can benefit from the final study to predict a better demand and supply, in addition the following benefits are also crucial for EPSA

- The experts who forecast the demand can access the system, analysis different patterns of the distributed drug to make appropriate decisions on end user demand request.
- To increase the forecasting accuracy through adopting the demand forecasting algorithm.
- Can also forecast the future demand and a better supply decision.
- Can take the advantage of efficient resource utilization and lead time management.
- Can significantly manage Bullwhips effect in the supply chain.
- Other researchers may use this study as their reference document for their research in related study.

## 1.8. Operational Definition

**The maximum months of stock** is the largest amount of each pharmaceutical avail in the warehouse. Having more than the maximum in the warehouse, it is overstocked and risks having stocks expire before they are used.

**The minimum months of stock** is the level of stock at which actions to replenish inventory should occur under normal conditions.

**The emergency order point** is the level where the risk of stocking out is likely, and an emergency order should be placed immediately.

**Consumption**: Calculate the total amount of pharmaceuticals Issued out of the Pharmacy Store/hospitals/branch

**Days Out of Stock:** The total number of days a product was out of stock at warehouse.

**Stock on Hand:** If Months of Stock on Hand is less than 0.25, an emergency order is needed.

**Vital (The concern of this study):**

- Potentially lifesaving, the best tool to curl the morbidity of the area: in the absence of
- this pharmaceutical, the patient may die, or may be hurt.
- Crucial to provide the basic health services; without which it is impossible to deliver
- the basic services in a specific area
- It is mandatory 24 hours of a day, 7 days of a week, and or 12 months of a year
- Known, easily manageable by the health care team and is the most fit with the clinical setup of the facility than the other alternatives.

**Essential:**

- Effective against less severe but significant illness,
- In the absence of these items, it may be difficult to deliver the service; somehow one may deliver the service by using alternatives
- Are manageable by the medical staff and fits in a better way to the clinical setup of the facility than the other alternatives

**None Essential/Less Essential:**

- Effective for minor illnesses
- But may have high cost compared to its therapeutic advantage

## 1.9. Structure of the Study

This study composed of six chapters, In chapter-1 the introduction and background of the study, Statement of the problem, Motivation, Significance, Scope and Limitation, and Ethical considerations are illustrated briefly, in chapter-2-, the

literature and related works of the study are reviewed, in chapter-3, the research methodology, data understanding, data analysis techniques, variables, data pre-processing methods, data split and research method is discussed, in chapter-4 the result of the study, evaluation and discussion is briefly stated, and in chapter-5, the conclusion, recommendations and future work is mentioned.

# CHAPTER TWO: LITERATURE REVIEW

In this chapter we focus on Machine learning Algorithm, Predictive Modeling, Basics of Artificial Neural Network, Types of Artificial Neural Network, Working Principle of Artificial Neural Network, Basics of LSTM, Working Principle of LSTM, Basics of Gradient Descent, Working Principle of Gradient Descent, Regression models, Types of Regression, EPSA supply Chain System.

## 2.1. Predictive Modeling

Many predictive modeling techniques, including neural networks (NNs), clustering, support vector machines (SVMs), Gradient Descent and association rules, exist to help translate this data into insight and value. They do that by learning patterns hidden in large volumes of historical data. When learning is completed, the result is a predictive model. They are being applied to a myriad of problems such as recommender systems, the prevention of diseases and accidents fraud and abuse detection. The world to extract value from historical data obtained from people and sensors. People data includes structured customer transactions (for example, from online purchases) or unstructured data obtained from social media. Sensor data, on the other hand, comes from a barrage of devices used to monitor roads, bridges, buildings, machinery, the electric grid, and the atmosphere and climate. Predictive modeling techniques allow for the building of accurate predictive models, as long as enough data exists and data quality is not a concern [13].

### 2.1.1. Machine Learning Algorithm

Machine learning is the most fastest growing areas of computer science with far-reaching applications. From the data can be learn and detect meaningful pattern from the data. Machine learning become common tool almost any task that needs information extraction from large dataset [14].

Common machine learning algorithm types include:

**Supervised learning –** which is generate a function that maps inputs to desired output. classification problem is the major tasks of supervised learning: the learner

is required to learn (to approximate the behavior of) a function which maps a vector into one of several classes by looking at several input-output examples of the function.

**Unsupervised learning -** which models a set of inputs: labeled examples are not available.

**Semi-supervised learning -** which combines both labeled and unlabeled examples to generate an appropriate function or classifier.

**Reinforcement learning -** where the algorithm learns a policy of how to act given an observation of the world. Every action has some impact in the environment, and the environment provides feedback that guides the learning algorithm.

**Transduction -** similar to supervised learning, but does not explicitly construct a function: instead, tries to predict new outputs based on training inputs, training outputs, and new inputs.

**Learning to learn -** where the algorithm learns its own inductive bias based on previous experience.

Machine learning is about designing algorithms that allow a computer to learn. Learning is not necessarily involving consciousness but learning is a matter of finding statistical regularities or other patterns in the data [15].

### 2.1.2. Time Series Forecasting Methods

Moving average method, exponential smoothing Brown, 1956, Auto-Regressive Moving Average Huang and Yang, 1995, Auto-Regressive Integrated Moving Average Box and Jenkins, 1976, Random Walk model, time series decomposition and Z-Chart are univariate linear time series methods. The simple moving average is a series of arithmetic means and is applicable data present no trends. The exponential smoothing is an averaging method that reacts more strongly to recent changes in demand by assigning weights. Exponential smoothing methods have been developed in order to take into account trend and seasonality. Auto-Regressive Moving Average (ARMA) is appropriate for non-stationary data, when a system is a function of a series of unobserved shocks as well as its own behavior. Auto-Regressive Integrated Moving Average (ARIMA) is a more complex method that handles trend and seasonality, but requires larger data sets. An important issue

of time series analysis is the stationarity of the data. A stationary time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are constant over time. Most statistical forecasting methods are based on the assumption that the time series are stationary or can become stationary using mathematical transformations [8]

Time series forecasting is difficult, it is difficult even for recurrent neural networks with their inherent ability to learn sequentially. A time series is a set of evenly spaced, continuous, numerical data obtained at regular time periods. In the time series forecasting methods, the forecast is based only on past values and assumes that factors that influence the past and the present was continue influence the future. If future values of a time series can be predicted from its past values, then the series is deterministic. If the future of a time series can only be partly determined by past values, then the time series is stochastic or random [13].

### 2.1.3. Basics of Artificial Neural Network

Artificial neural networks (ANNs) were identified as potentially suitable for supply chain forecasting in the 1990's Leung, 1995 and have since been found to be very effective for developing forecasts. The ANNs architecture known as multi-layer perceptron (MLP) with back propagation is commonly used Beccali 2004 models. They suggest that a modified ANN can provide a better result. The modification proposed by Chang 2011 involves incorporating genetic algorithms (GAs) into the ANN as an improvement to define the ANN prior to training. GAs, first introduced by John Holland in 1975, are sometimes used to configure the topography of ANNs Kaylani 2010 [4].

Artificial Neural Networks are relatively crude electronic models based on the neural structure of the brain. The brain basically learns from experience. It is natural proof that some problems that are beyond the scope of current computers are indeed solvable by small energy efficient packages. This brain modelling also promises a less technical way to develop machine solutions. This new approach to computing also provides a more graceful degradation during system overload than its more traditional counterparts. Artificial Neural Network (ANN) is preferred as a superior forecasting model because it addresses the limitations of time series

models by efficient non-linear mapping between input and output data. For forecasting purpose, ANN neither requires any statistical information nor stationary nature of data series. It is basically a black-box model, data driven and a function approximate to solve complex problem. ANN has the ability of self-learning, self-organizing and self-adapting to the data series. Therefore, ANN is applied in various fields of engineering and management including supply chain for prediction purpose when mapping relation between input and output is not known at priori. ANN being a data driven method, it simply maps the input and output without emphasizing on the pattern of the data series [16].

### 2.1.4. Types of Topologies in Artificial Neural Network

Individual artificial neurons are interconnected is called topology, architecture or graph of an artificial neural network and interconnection of topologies can be done in numerous ways results in numerous possible topologies that are divided int two basic classes. Fig. 1 shows the two topologies; the left side of the figure represent simple feedforward topology (acyclic graph) where information flows from inputs to outputs in only one direction and the right side of the figure represent simple recurrent topology (semi cyclic graph) where some of the information flows not only in one direction from input to output but also in opposite direction [16].

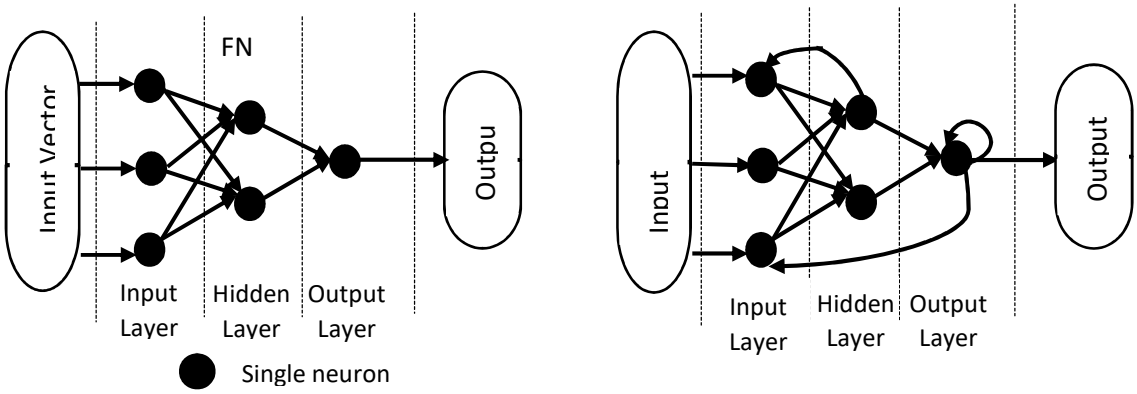



**Figure 2. 1. Feed-Forward (FNN) and Recurrent (RNN) Topology of an Artificial Neural Network [17]**

Feed Forward- Feedback is a type of connection where the output of one-layer routes back to the input of a previous layer, or to same layer information must flow from input to output in only one direction with no back-loops [16].

There are no limitations on number of layers, type of transfer function used in individual artificial neuron or number of connections between individual artificial neurons.

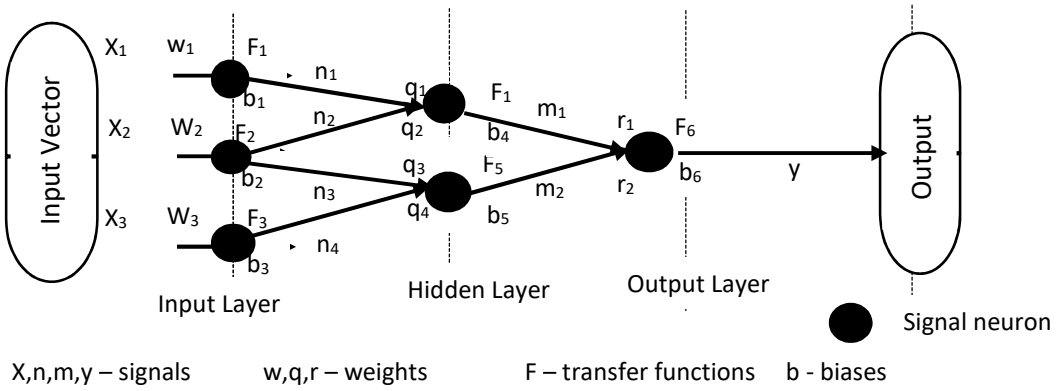



**Figure 2. 2. Feed-Forward Artificial Neural Network [16].**

Artificial neural network with the recurrent topology is called recurrent artificial neural network. It is similar to feed-forward neural network with no limitations regarding back loops. In these cases, information is no longer transmitted only in one direction but it is also transmitted backwards. This creates an internal state of the network which allows it to exhibit dynamic temporal behavior. Recurrent artificial neural networks can use their internal memory to process any sequence of inputs. The most basic topology of recurrent artificial neural network is fully recurrent artificial network where every basic building block (artificial neuron) is directly connected to every other basic building block in all direction. [17] [18].

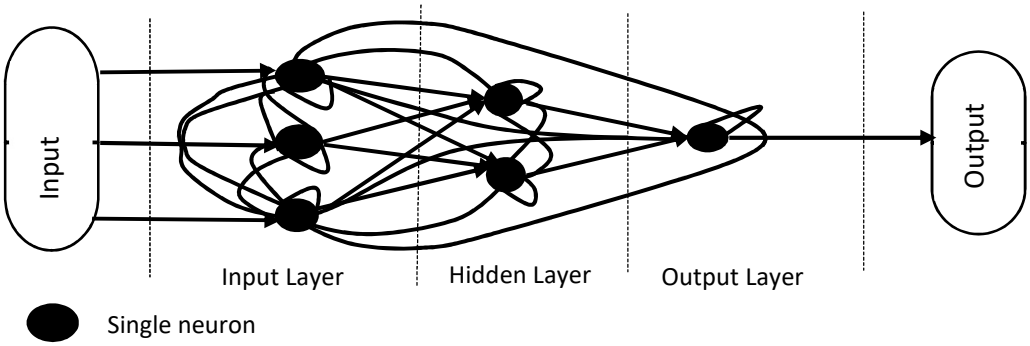



**Figure 2. 3. Fully Recurrent Artificial Neural Network [16].**

### 2.1.5. Working Principle of Artificial Neural Network

Artificial neural networks (ANNs) are biologically inspired computer programs designed to simulate the way in which the human brain processes information. ANNs gather their knowledge by detecting the patterns and relationships in data and learn (or are trained) through experience, not from programming. Artificial neurons or processing elements (PE), connected with coefficients (weights), which constitute the neural structure and are organized in layers. Each PE has weighted inputs, transfer function and one output. The behavior of a neural network is determined by the transfer functions of its neurons, by the learning rule, and by the architecture itself. The weights are the adjustable parameters and, in that sense, a neural network is a parameterized system. The weighed sum of the inputs constitutes the activation of the neuron. The activation signal is passed through transfer function to produce a single output of the neuron. Transfer function introduces non-linearity to the network. Training the inter unit connections are optimized until the error in predictions is minimized and the network reaches the specified level of accuracy. Once the network is trained and tested it can be given new input information to predict the output [17].

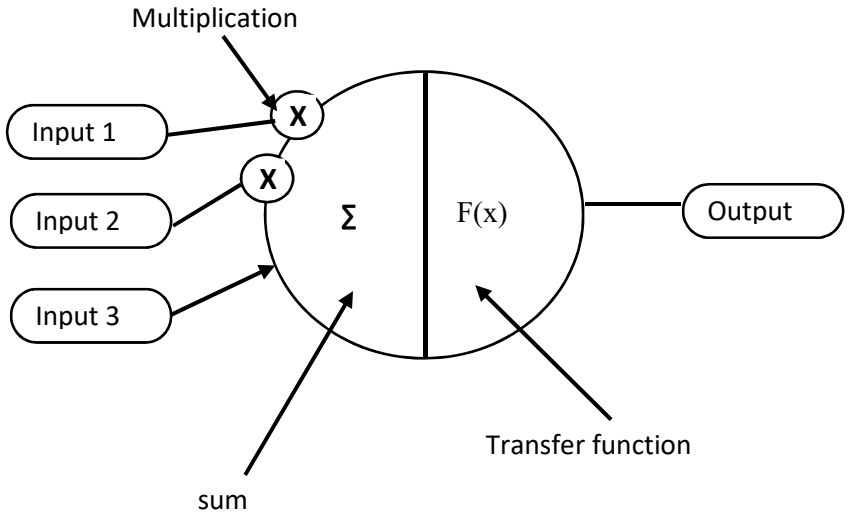



**Figure 2. 4. Artificial Neural Network design [19].**

BP trains multilayer feedforward networks with differentiable transfer functions to perform function approximation, pattern association and pattern classification. It is the process by which the derivatives of network error, with respect to network

weights and biases, are computed to perform computations backwards through the network. Computations are derived using the chain rule of calculus. There are several different BP training algorithms with a variety of different computation and storage requirements. No single algorithm is best suited to all the problems. All the algorithms use the gradient of the performance function to determine how to adjust the weights to minimize the performance. For example, the performance function of feed forward networks is the Mean Square Error (MSE). The basic BP algorithm adjusts the weights in the steepest descent direction (negative of the gradient); that is, the direction in which the performance function decreases most rapidly. The training process requires a set of examples of proper network inputs and target outputs. During the training, the weights and biases of the network are iteratively adjusted to minimize the network performance function [16].

### 2.1.6. Basics of Long Shor-Term Memory (LSTM)

LSTM is a recurrent neural network (RNN) architecture that had been shown to outperform traditional RNNs on numerous temporal processing tasks Time series problems are conceptually simpler than many tasks already solved by LSTM. They often do not require RNNs at all, because all relevant information about the next event is conveyed by a few recent events contained within a smalltime window. LSTM is run as a pure" autoregressive (AR) model that can only access input from the current time-step, reading one input at a time, while its competitors [19].

Recurrent neural networks are well suited to supervised learning problems where the dataset has a sequential nature. RNNs are essentially neural networks with memory. They can remember things from the past, which is obviously useful for predicting time-dependent targets. Applying them to time series forecasting is not a trivial task. The main idea behind recurrent neural networks is using not only the input data, but also the previous outputs for making the current prediction. This idea makes a lot sense we could build neural networks passing values forward in time. However, such simple solutions usually do not work as expected. They are hard to train and they are forgetful. Rather, there must be a system with some kind of memory. There are two popular and efficient RNN models that work really well: LSTM and gated recurrent unit [20].

LSTM Hoch Reiter and Schmid Huber 1997 are a gated memory unit for neural networks. It has 3 gates that manage the contents of the memory. These gates are simple logistic functions of weighted sums, where the weights might be learnt by backpropagation. It means that, even though it seems a bit complicated, the LSTM perfectly gets into the neural network and its training process. It can learn what it needs to learn, remember what it needs to remember, and recall what it needs to recall, without any special training [21].

### 2.1.7. Working Principle of Long Shor-Term Memory (LSTM)

Recurrent nets are a type of artificial neural network designed to recognize patterns in sequences of data, such as text, genomes, handwriting, the spoken word, or numerical times series data emanating from sensors, stock markets and government agencies. These algorithms take time and sequence into account, they have a temporal dimension [19].

The LSTM network, or LSTM network, is a recurrent neural network that is trained using Backpropagation Through Time and overcomes the vanishing gradient problem.

Create large recurrent networks that in turn can be used to address difficult sequence problems in machine learning and achieve state-of-the-art results. Instead of neurons, LSTM networks have memory blocks that are connected through layers. A block has components that make it smarter than a classical neuron and a memory for recent sequences. A block contains gates that manage the block's state and output. A block operates upon an input sequence and each gate within a block uses the sigmoid activation units to control whether they are triggered or not, making the change of state and addition of information flowing through the block conditional.

There are three types of gates within a unit:

- **Forget Gate**: conditionally decides what information to throw away from the block.
- **Input Gate**: conditionally decides which values from the input to update the memory state.

- **Output Gate**: conditionally decides what to output based on input and the memory of the block.

The unit is like a mini-state machine where the gates of the units have weights that are learned during the training procedure [22].

## 2.2. Regression Models

In [23] A model with a single independent variable is called a simple regression model. Multiple regression refers to a model with on dependent and two or more independent variables. A successful regression analysis provides useful estimates of how previous changes in each of the independent variables have affected the dependent variable. As a forecasting approach, regression analysis has the potential to provide not only demand forecasts of the dependent variable but useful managerial information for adapting to the forces and events that cause the dependent variable to change.

In [24] The regression analysis is a statistical tool that aims to explore the strength of the relationship between one dependent variable which is the respond variable and one or many other changing variables known as the independent variables or the explanatory variables. It is a powerful technique that uses the values from historical data of one or more variables to develop a model that helps in predicting the value of the dependent variable.

### 2.2.1. Types of Regression Analysis

[24] The regression analysis has different types according to the nature of the relationship between the dependent and the intendants variable. The main two types of regression analysis according to the relationship between the dependent and the in-Descent variable are Linear regression analysis and Nonlinear Regression analysis.

**The Non-linear regression: -** a regression in which the dependent variables are modeled a non-linear function of model parameter and one or more in Descent variables.

The linear regression analysis has two main types according to the number of predictors.

Simple regression analysis: - this type of regression assesses the relationship between one dependent variable and only and independent variable. It simply has an equation of a line.

Where Y is the value of the dependent variable to be predicted or explained

Alpha (a) is a constant that equals the value of Y when the value of X=0.

Beta β or is the coefficient of X. It is the slope of the regression line that can explain how much Y changes for each one-unit change in X.

X is the value of the Independent variable (X), which is the factor that predicting or explaining the value of Y.

E is the error in predicting the value of Y.

**Multiple regression analysis: -** This type of regression assesses the relationship between one dependent variable and many in Descent factors. The form of this regression model fore given (k) in Descent variables.

## 2.3. EPSA Supply Chain System and IPLS

Ethiopian Pharmaceutical Supply Agency (EPSA) designed to make center of excellence at all aspects of the agency provide the service to the public and on its working area. The agency working with different sectors to use the maximum effort and significantly to improving quality assured pharmaceuticals at an affordable price to the public and sustainable availability. The agency plan to reduce national pharmaceutical wastage and improve distribution pharmaceuticals to service delivery points. The Agency has developed and further revised standard operating procedures (SOP) for Integrated Pharmaceuticals Logistics System (IPLS). This SOP manual encompasses major areas of pharmaceutical management at health facilities including; Logistics Management Information System, inventory control system, recording and storage. IPLS is the term applied to the single pharmaceuticals reporting and distribution system based on the

overall mandate and scope of the EPSA. It aims to ensure that patients always get pharmaceuticals they need. To be successful, the system must fulfil the six rights of supply chain management by ensuring the right products, in the right quantity, of the right quality, at the right place, at the right time and for the right cost. The IPLS integrates the management of essential pharmaceuticals drugs purchased. It is the primary mechanism through which all public health facilities obtain essential and vital pharmaceuticals. Products included on the National pharmaceuticals procurement List (NPPL) are supplied and managed through the IPLS. One of the first concrete steps to move the integrated system from concept to detailed implementation step was the development of the Standard Operating Procedures (SOP) Manual for health facilities of Ethiopia [9].

## 2.4. Related Work

In this section, related works related to the study was reviewed and discussed. One of our objectives is to use the variety of data available in the EPSA to develop predictive model to forecast demand using machine learning tools. In order to analyze this problem better, focus on literature review.

In [7] this study conducted to determine the accuracy of statistical forecasting techniques used in inventory planning of Apollo Pharmacy of the Apollo group in an Indian community at brick level. The gap analysis was carried out and it was understood that information from customers to supplier is not captured. According to preliminary surveys, forecasting does not factor in their SCM process and there is underestimation of the importance of forecasting as a part of SCM. This article demonstrates the application and screening of simple forecast models (Moving Average, Exponential smoothing, Winter's Exponential) to forecast demand for pharmaceuticals. Two products were empirically chosen for the study Okacet 10mg tablet (seasonal demand, contains Citrizine) and Stamlo Beta tablet (non-seasonal, contains Amlodipine 5 mg & Atenolol 50 mg). Accuracy assessment parameters indicate that the least sales forecast error for Okacet (seasonal) was obtained using WES ($\alpha$=0.2, $\beta$ = 0.1, $\gamma$ = 0.01, ET = 92.28) and (MAPE = 27.50) indicated highest forecast accuracy, whereas for Stamlo Beta (non-seasonal) WES provided no additionally superior forecast as other models.

In [25] this study demand forecasting in pharmaceutical supply chains was conducted. In the case study, demand forecasting experiments have been performed for a specific pharmaceutical product ACT0002UZ01. The historical weekly sales product data contains 41 data points. Three experimental scenarios based on application of different forecasting methods are investigated applying the simple moving average method, multiple linear regression, and symbolic regression with genetic programming. For regression scenarios, the following factors potentially affecting the product demand have been taken into consideration: a distributor pricelist; the discounted selling price of the product; a week number of sales in a month; and weekly average currency rate. For each

scenario, forecasting output and forecast errors are analyzed and applications feasibility and implications are provided. The results of experimental analysis of three forecasting scenarios show that symbolic regression-based forecasting model provides the best fitting curve to history demand data, lower error estimates across all scenarios and performed experiments, and the ability to more accurately predict demand peak sales in the study.

[26] The purpose of the main survey was to investigate the influence of the economic crisis to the logistics services sector in Greece. They introduce time series methods, which was suitable for short term predictions. Error was measured using mean square error. The data was time series starting from 2009-2011. The dataset was consumption and purchase of the drug from the cardiology department indicated for the thrombolytic treatment of suspected strokes. Augmented Dickey-Fuller test was applied for checking stationarity of the dataset. In order to decide whether there is a cyclic component in their data, they use the seasonal subseries plots which are a tool for detecting seasonality in a time series. After examining the dataset comparing the three models are conducted: moving average, Auto Regressive and seasonal exponential smoothing model. More advanced forecast methods cannot be used when the available data are so few. The findings of this research were simple seasonal exponential smoothing model is fit with the data and this model is appropriate for series with no trend and a seasonal effect that is constant over time. On this paper error measure techniques were not stated.

[27] this study focusses on to predict different drug consumption. Comparison was performed on decision trees regression and artificial neural networks. This study follows empirical study. 3-year dataset with 407000 records of a drug distribution center of Tehran capital of Iran was used. The data partitioned randomly into two subsets 70% of the records was for training and the remainder was for testing. Root-mean-square deviation (RMSE) and normalized RMSE (NRMSE) considered as the comparison scales. The study identified using data mining techniques, hidden relationships between the variables. In this study data mining techniques such as a combination of association rule and prediction algorithms have been found helpful in forecasting. The findings of this study are with

existence variables and data mining techniques are more efficient than statistical ones. ANN and DT algorithms are applied. After an empirical study, the methods are stable. DT would perform a little better than ANN in most of these cases.

In [28] presented in this study demand forecasting in food company using time series approach. The past data was used to develop several autoregressive integrated moving average models by suing BOX Jenkins time series procedure and the performance of the model was measured based on four performance criteria Akaike criterion, Schwarz Bayesian criterion, maximum, likelihood, and standard error and the model was selected. Data comes from Moroccan food manufacturing starting from 2010 until December 2015. The findings of this study were the model can be used for modeling and forecasting the future demand in manufacturing food industry.

[29], according to this article aims to reduce the errors in prediction demand and stock out, inventory health, man power and logistical resources. Firstly, identify several causal features and use a combination of feature embeddings, MLP and LSTM. The experiment conducted on a dataset of a year's and then thousands of products from Flipkart. Multi-layer perceptron (MLP) and a long-short term memory (LSTM) network was combined for handling of associative and time-series features, and a mixture density network at the output to flexibly capture output uncertainty. Therefore, the final model, AR-MDN, is associative, recurrent and probabilistic. The comparisons with existing methods and select the best of the breed non-deep learning-based method and compare with that. Next, the comparison with various architectures among deep learning models done. The results of this study show AR-MDN model is the best.

[3], This paper states to fulfill the overestimated demand, excess production results in extra stock keeping which ties up excess inventory. The author used a sales data. The researcher was developed demand forecasting technique which is modeled by artificial intelligence approaches using artificial neural networks. MATLAB 7.0 was used for ANN simulation. According to this study the result indicates Train LM method performs more effectively than the other tanning method and the more reliable forecast for the case. The methodology can be considered as a successful

decision support tool in forecasting. The ability to increase forecasting accuracy was result.

[30], This paper focuses on a various demand forecasting models to predict product demand for grocery items using Python's deep learning library. The data source for the study was Kaggle competition which aims to forecast demand for millions of items at a store and day level for a South American grocery chain. Performance comparison was done with different open source modeling techniques to predict a time dependent demand at as store sku level. These demand forecasting models were developed using Keras and scikit-learn packages and the comparisons along the following dimensions: predictive performance, runtime, scalability and ease of use were done. The forecasting models explored in this study are Gradient Boosting, Factorization Machines, and three variations of Deep Neural Networks (DNN). Performance measured by Normalized weighted mean square log error (NWMSLE) for selection the best model. After comparing with other mode, they decided to use the output of this model as inputs to the neural network approach to improve the overall accuracy of the model.

[31], focus of the study by using hybrid methodologies to forecast selling amount of the product. To forecast the demand of warehouses four different algorithms are determined. The Selected algorithms are Bayesian Network, Multi-Layer Perceptron, Support Vector Machine and Linear Regression. Stacked generalization methodology was used for combining four different machine learning models in the study. The experimental results show that the selected approach achieves better results for forecasting demands of warehouses. In terms of MAPE, the proposed method provides nearly 5% less error than best result of stand-alone models. Moreover, error rate nearly 9% compared to Stacked Generalization.

[4], Forecast based on demand behaviors of segments of customers was the focus point of the study. Four years of historical data and nearly one million observations of delivery events, including customer, distribution center, quantity, and product type usage and the effects of external factors by using data mining techniques and to develop a prediction method accurate forecast demand. Different clustering

techniques used, such as K-means, self-organizing maps (SOMs), and fuzzy clustering were explored after removal of outliers, the data was normalized and then re-segmented with k-means. Customers were segmented using tools R program and using k-means algorithm. After monthly segmented customers, forecasting models were applied to assess the viability of the segments. This paper has presented a method for transposing a high number of individual customers into a small number of clusters of customers with similar demand behavior. Finally, this paper discussed that the presence of outlier's data had an influence on the result. The finding of this study enables the supplier BM to understanding the change on demand. The change was due to seasonal effect and whether capacity changes are necessary.

[32] The research approach is robust and efficient time series model which can be applied to linear, non-linear, stationary or non-stationary data series. They pre-processed past demand series through DWT to obtain different demand sub series for capturing the pattern inherent with data. The resulting sub-series are given as input to an ANN model for training and testing. The proposed approach validated with dataset from open literature and comparative study between DWT-ANN and ARIMA has been made. Findings of the study was by comparing with well-known time series model ARIM the forecasting accuracy of the DWTANN model by estimating the main square error (MSE). The MSE was comparatively less in case of DWT-ANN. The approach can reduce BWE and NSAmp. Net-stock was less varying as compared to demand variance in case DWT-ANN model is adopted.

[33] this research is discussed critically about the effect of demand forecasting. To monthly data of six different consumer products are used to assess the performance of the method. Starting from January 2009 to August 2011 was used. In order to study the pattern and trend of the data time series plots was plotted for each product. To measure the performance of the model, mean absolute percentage error (MAPE) was used. In the study two most popular neural network architectures are applied, multilayer perceptron (MLP) and radial basis function (RBF). Multilayer perceptron architecture might be suitable for the dataset. To build Broyden Fletcher-Gold-Shannon (BFGS) algorithm the model training

algorithm of the multilayer perceptron network was used. The forecasting model conducted based on support vector machine approach was the regression type 1 and the kernel was radial basis function. The results indicated that Support Vector Machine had a better forecast quality than Artificial Neural Network in every category of products. The Findings of the research was the margin difference of mean absolute percentage error from these two methods was significantly high when the data was highly correlated.

[34] Experimenting and comparing the performance of the Gradient tree Boosting (GTB) with other forecasting techniques and machine learning algorithm was the focus of the research. The datasets were computer equipment retailer sales dataset and Bike Sharing demand competition. The GTB was competitive in both datasets, it shows better result on the Bike dataset and also on computer retailer sales dataset GTB shows limited information. GTB or Neural Network can obtain better predictions than using the baseline naïve forecasting model. This paper not include how to they measure the error.

[6] ANN was deployed to forecast the demand and at the end perfect accuracy by modeling it mathematically. R software and MATLAB are used to create the neural network. The data is organized and results are compared using python. The analysis of the paper has been done using demand forecasting of American multinational retail corporation. The main problem of demand forecasting is modeled carefully and all factors that might affect the demand should be taken into viewpoint and should be independent of each other as much as possible. The results show that ANN's can be used on very large datasets as well and still produce very less errors. R provide better accuracy but it consumes more time and MATLAB, produce results faster in comparison than R and all better parameter optimization.

# CHAPTER THREE: METHODOLOGY

Accurate forecasting requires the existence of quality consumption data from the source and existence of simple forecasting tools is mandatory, in this study the overall research objective was to forecast pharmaceutical demand. This study was conducted in Ethiopia Pharmaceuticals Supply Agency based on Consumption dataset. Phyton and RStudio tools are utilized as means to address the research problem. It is essential to investigating whether there are similar researches and tools having similar functionality. In addition, papers written on demand forecasting was reviewed to get an understanding of the techniques, methods and algorithms of demand forecasting. Python tools are utilized as means to achieve the research problem.

Learning algorithms were used for the forecasting of demand and in EPSA dataset. LSTM, MLP and Auto ARIMA three algorithms were selected for prediction and the models were trained on purposely selected item and selected future from the original dataset. Later the performance of the models was evaluated using MSE, performance measures.

## 3.1. The Existing Data Collection Method of the Agency

The agency uses the quantification outputs for iterative process of reviewing and updating the quantification data and assumptions, and recalculating the total commodity requirements and costs to reflect actual service delivery and consumption; as well as changes in program policies and plans, over time.

Quantification is the process of estimating the quantities and costs of the products required for a specific health program (or service) and determining when the products should be delivered to ensure an uninterrupted supply for the program. Quantification is a critical supply chain activity that links information on services and commodities from the facility level with the program policies and plans at the national level; it is then used to inform higher-level decision-making on the financing and procurement of commodities. The results from a quantification can be used to help maximize the use of available resources for procurement; advocate for mobilization of additional resources, when needed; and inform manufacturer

production cycles and supplier shipment schedules. The results of a quantification should be reviewed and updated at least every six months, and more frequently for rapidly growing or changing programs. The Reviewing and Updating the Quantification section in this document provides more detail on monitoring and updating quantifications.

### 3.1.1. Purpose of the Quantification

It is important to identify the purpose of the quantification needs. Examples of the purpose of a quantification include the following:

- Provide data on specific commodity requirements and costs for the government's annual budget allocations.
- Inform donors about funding requirements and advocate for resource mobilization for commodity procurement.
- Estimate commodity needs and assess stock status of the in-country supply pipeline to identify and correct supply imbalances.
- Support an estimate of commodity procurement, storage, and distribution costs.

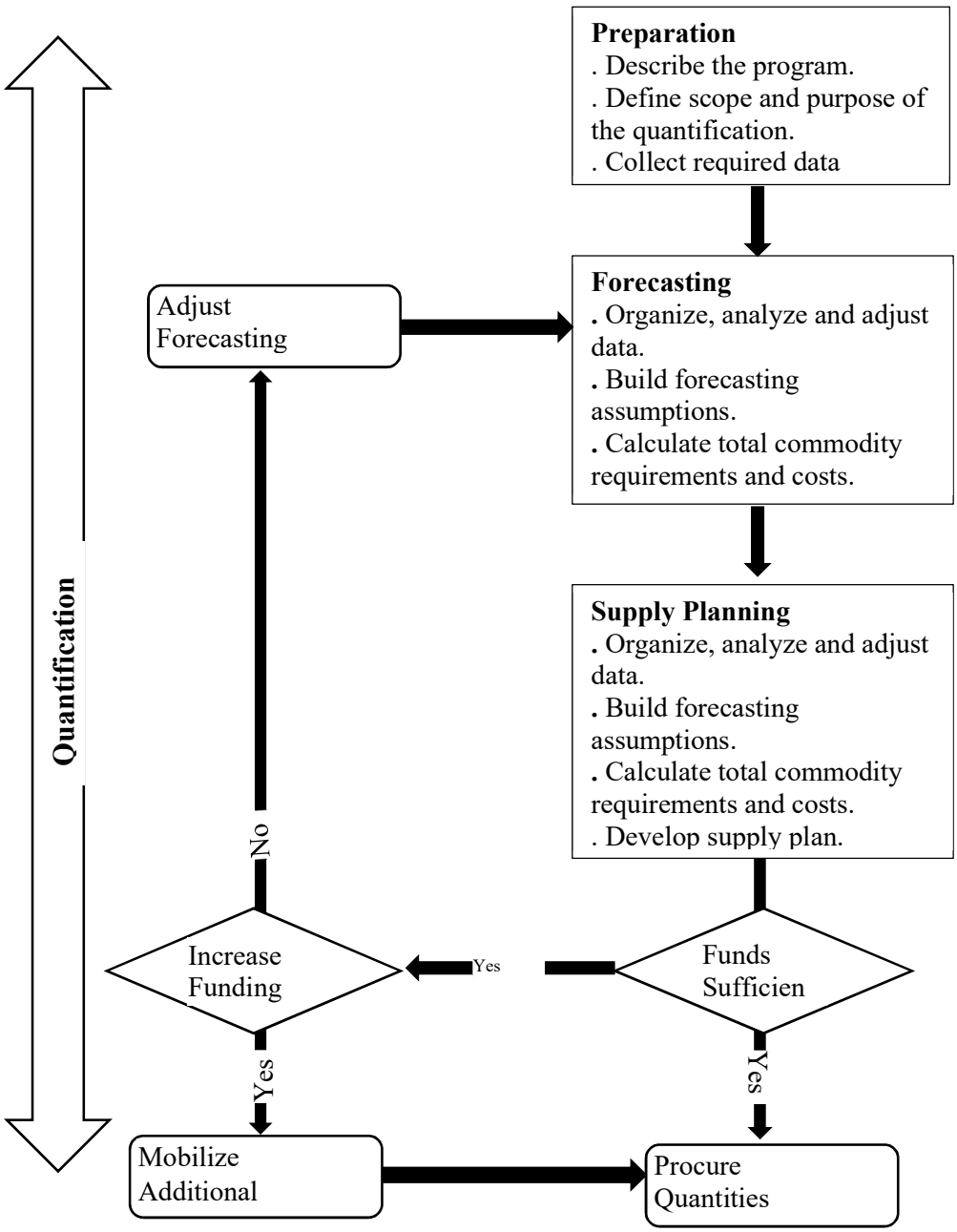



**Figure 3. 1. Steps in Quantification [12]**

This section describes a standardized, step-by-step approach to quantification. The three basic steps are preparation, forecasting, and supply planning.

### 3.1.2. Data for Forecasting

The agency uses consumption data for forecasting. Consumption data is quantity of each product dispensed or consumed during the past 12 months Consumption data are historical data on the actual quantities of a product that have been

dispensed to patients or consumed at SDPs, within a specified period, which can be reported monthly, bimonthly, or quarterly. Daily consumption data can be found in pharmacy dispensing registers or other point-of-service registers. Where a well-functioning LMIS captures and aggregates these data from SDPs, aggregated consumption data can be found in monthly facility level and annual program-level reports. When  tablets of co-trimoxazole were consumed. consumption data are used, the forecast is based on the quantities of products consumed in the past. Consumption data are most useful in mature, stable programs that have a full supply of products and reliable data.

Framework of the Methodology

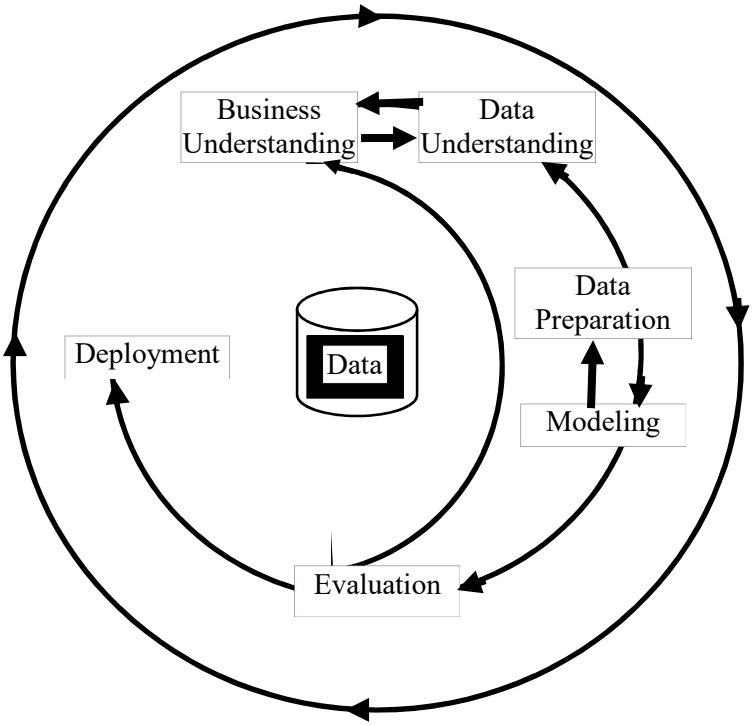



**Figure 3. 2. Framework of the Research Methodology**

## 3.2. The Proposed Data Collection Method

The agency has procurement list, it must be providing to its beneficiaries. The agency also selects about 1400 Items by agency professionals and other sectors as national level. Out of 1400 items 57 of them are vital of vital items. Therefore, the

scope of the study was only vital of vital items. 4 items are selected out of 57 by using purposive and stratified random sampling. The selected items are used as tracer in the agency also on their vitality, un replacement by another item, their frequency of distribution and also expert recommendation. Items selected for this study are Atropine Sulphate - 1mg/ml - Injection (Sulphate), Insulin Isophane Biphasic (Soluble/Isophane Mixture) - (30 + 70) IU/ml in 10ml Vial - Injection (Suspension), Dextrose - 40% in 20ml - Iv Infusion, Tetanus Antitoxin (TAT), Equine - 1,500 IU/ml in 1ml Ampoule – Injection. We used dataset starting from 2016 to 2019 because the data base before 2016 not include the consumption dataset. About 30,038 records are collected from the HCMIS database. Among 8 features are removed out of 12 because to make the dataset suitable for the algorithms. For the implementation of the research real consumption data are extracted and used from the HCMIS database.

## 3.3. Data Preparation

In [26] [27] time series forecasting research includes Date and Quantity are selected features. The study used dataset starting from 2016 to 2019. About 30,038 records and about 12 variables which contains information on health pharmaceuticals and the data collected from the HCMIS database. Consumption data of pharmaceuticals product related features such as Item Name, Unit, Received, Consumption, AMC, Min, Max, SOH, Amount, Month, Status. The study prepares the dataset for time series forecasting operation. The data contain a dependent features consumption indicating the consumption of the products. Therefore, the study selected date and consumption to build the time series model from the original dataset. The selected data is not in a format that is suitable for our work therefore, the data transformed from xlsx format to csv excels format.

**Table 3. 1. Features of the Dataset**

| Features | Data type | Description |
|---|---|---|
| Item | String | Indicates the name of all item that are distributed for all branches for each year. This variable contains text value. |
| Unit | String and int | Indicates the unit of item. It helps to identify item that has the same name but they have different unit. Contains numeric value and text value. |
| Received | int | This variable shows that the amount of Item received in the central warehouse and contains int value |
| Consumption | int | Calculate the total amount of pharmaceuticals Issued out of the Pharmacy Store using the beginning balance in the store, Quantity Received, Loss/Adjustment and Ending balance in the store. |
| AMC | int | Average Monthly consumption of central warehouse and contains numerical values. |
| MOS | int | Month of Stock |
| Min | int | The variable that shows the level of stock. |
| Max | int | Is the largest amount of each pharmaceutical a center should hold at any one time. |
| SOH | int | The variable shows that the current available stock on Hand and contains numerical values. |
| Amount | int | Indicates the price of an item and contains numerical value. |
| Month | Date | This variable show that the amount of consumption on the specific month |
| Status | string | The variable shows that the status of the stock in the warehouse and contains string value. |

**Table 3. 2.The Independent Features**

| Features | Data type | Description |
|---|---|---|
| S/N | int | The count number |
| Item | String | Indicates the name of all item that are distributed for all branches for each year. This variable contains text value. |
| Unit | String and int | Indicates the unit of item. It helps to identify item that has the same name but they have different unit. Contains numeric value and text value. |
| Consumption | int | Calculate the total amount of pharmaceuticals Issued out of the Pharmacy Store using the beginning balance in the store, Quantity Received, Loss/Adjustment and Ending balance in the store. |

**Table 3. 3. The Dependent Features**

| Features | Description |
|---|---|
| Consumption | Calculate the total amount of pharmaceuticals Issued out of the Pharmacy Store using the beginning balance in the store, Quantity Received, Loss/Adjustment and Ending balance in the store. |

## 3.4. Data Preprocessing

Data preprocessing is the basic stage in order to preprocess and prepared the data in the way it will be suitable to feed for the algorithms. In this stage, data cleaning, missing value handling, and data splitting has been performed.

Data preprocessing phase included four steps: formatting, cleaning, replacing missing values, purifying data to eliminate dataset errors and merging data in order to reduce their volume. In purifying the data, according to the manual data entrance process, there is a probability of data error. In this dataset, some error where found and corrected, such as Item names are written in different year in different spalling, with space and without space. In order to solve all this problem, we need to write in standard form. Algorithms learn from data. It is critical having the right data for the problem we want to solve. Even if we have good data, we need to make sure that it is in a useful scale, format and even that meaningful features are included. After we have selected the data, we considered how we are going to use the data. For this study the preprocessing steps was done in the following pattern:

### 3.4.1. Cleaning

During collecting, entering data, transforming, extracting data, exploring, analyzing data and submitting the draft report for peer review the data may have some errors, outliers and some missing. Therefore; data cleaning was needed for fixing an error occurred [35]. Observation that contains high values were deleted, and on the study. Clearing is done to spot out errors and fix them.

**Standardizing Item names**

| Item |
| --- |
| Tetanus Antitoxin (Human) Equine - 1500 Units - Injection |

| Mapped Item |
| --- |
| Tetanus Antitoxin (TAT), Equine - 1,500 IU/ml in 1ml Ampoule - Injection |

The same item written in different spalling, character and spacing to solve such kind of problems we mapped the item name with correct one. About 80 rows was written wrongly from 108 consumption for specific item and corrected in this study.

### 3.4.2. Searching for Outlier

A boxplot is a graph that gives us a good indication of how the values in the data are spread out. In order to get the advantage of taking up less space, outliers detect

and removing and comparing distributions between many groups or datasets, a boxplot and time series was plotted to visualize on for the study. After visualizing the existence of the outlier, the study removed the Interquartile range of the data set. To find the IQR the middle 50% of values. IQR is the difference between the median (middle value) of the lower(Q1) and the upper half of the data(Q3). Removing and Replacing the outlier are the common features on handling the outliers.

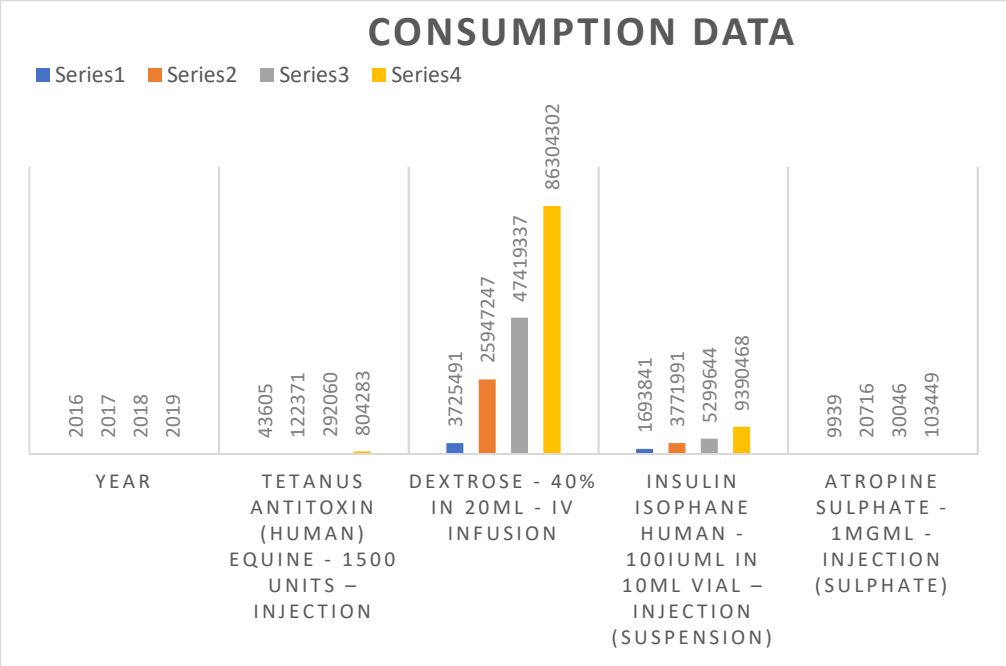



**Figure 3. 3 Items from 2016 to 2019 Annual Consumption**

The above data illustrates an aggregated data for the selected items four years which is extracted from the consumption dataset, and this data consider and ready for further steps.

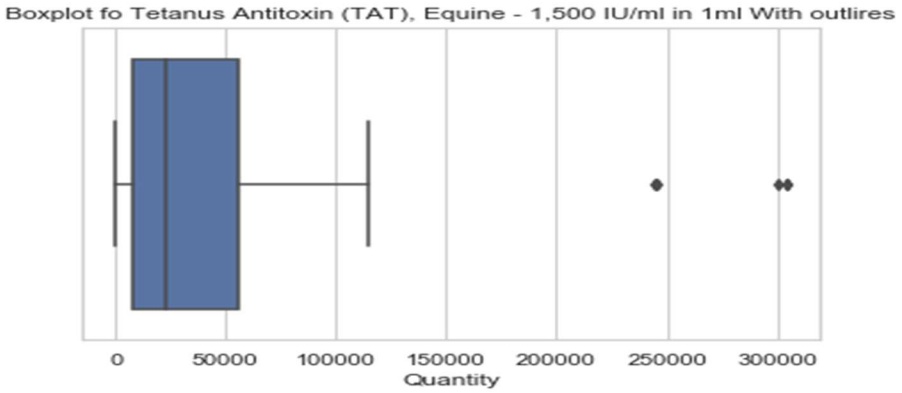



**Figure 3. 4. Boxplot for Tetanus Antitoxin (TAT), Equine - 1,500 IU/ml in 1ml**

During the study as on the boxplot above indicated three outliers are detected, removed and replaced by imputation techniques according to figure 3.4 shows.

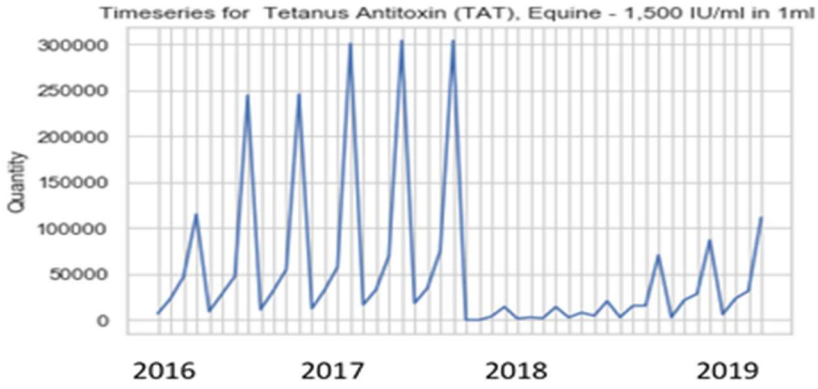



**Figure 3. 5. Timeseries for Tetanus Antitoxin (TAT), Equine - 1,500 IU/ml in 1ml**

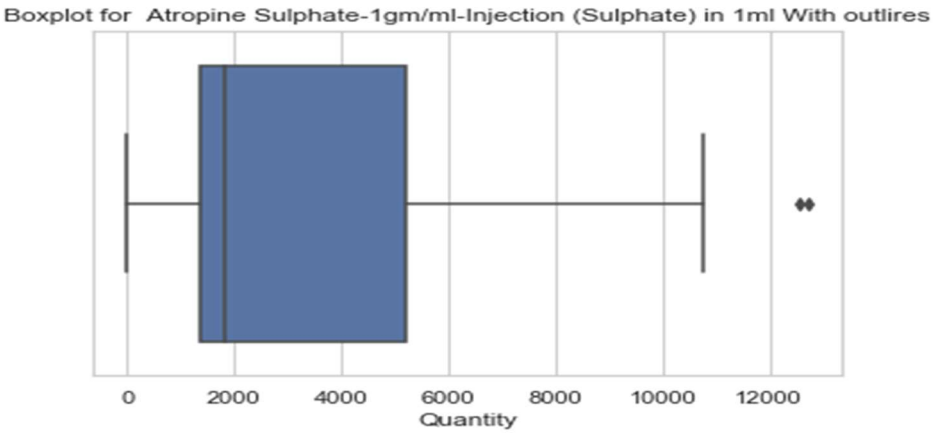



**Figure 3. 6. Boxplot for Atropine Sulphate-1gm/ml-Injection (Sulphate)**

During the study as on the boxplot above indicated two outliers are detected and removed and replaced by imputation techniques according to figure 3.6 shows.

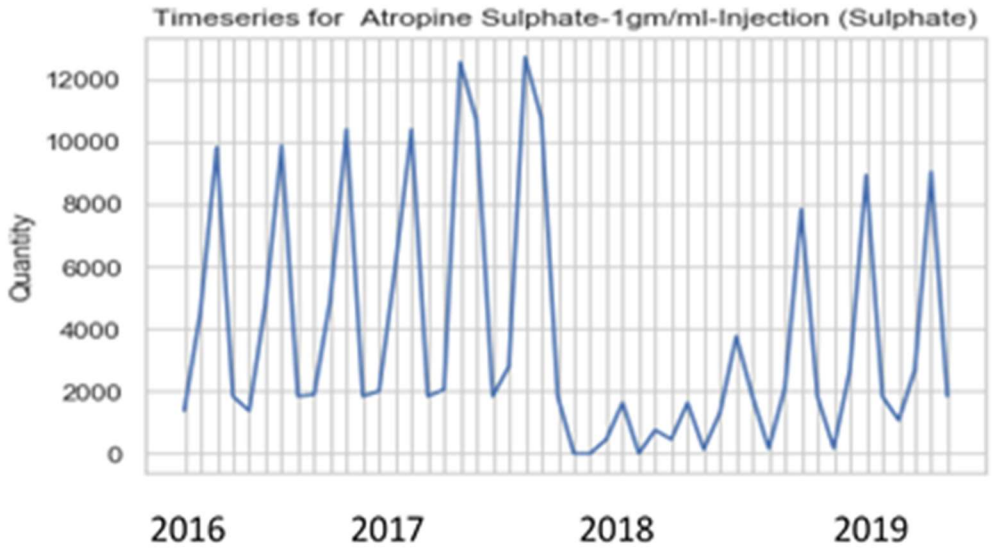



**Figure 3. 7. Time Series Atropine Sulphate-1gm/ml-Injection (Sulphate)**

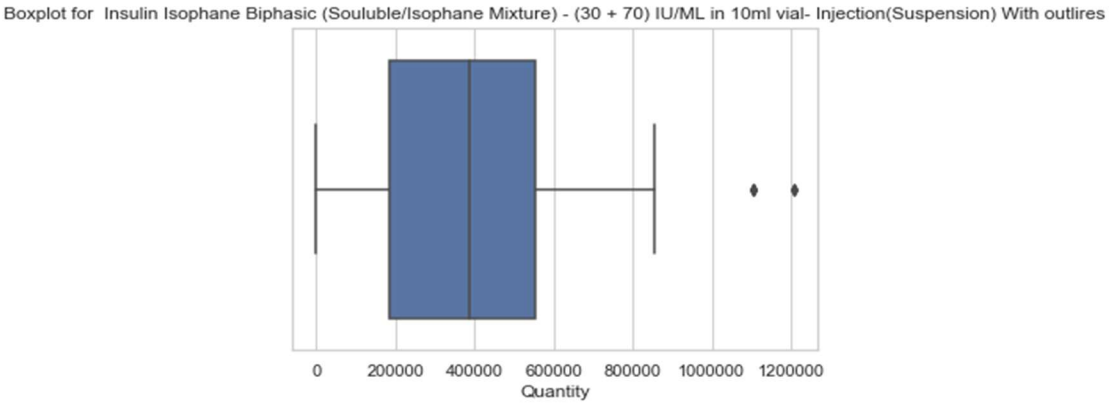



**Figure 3. 8. Boxplot for Insulin Isophane Biphasic (Soluble/Isophane Mixture) - (30 + 70) IU/ml in 10ml Vial - Injection (Suspension).**

During the study as on the boxplot above indicated two outliers are detected and removed and replaced by imputation techniques according to figure 3.8 shows.

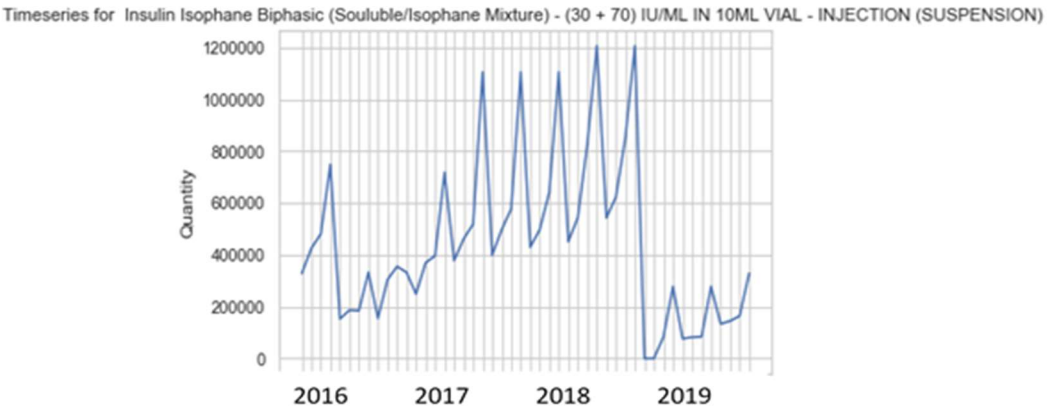



**Figure 3. 9.** Timeseries FOR Insulin Isophane Biphasic (Soluble/Isophane Mixture) - (30 + 70) IU/ml in 10ml Vial - Injection (Suspension)

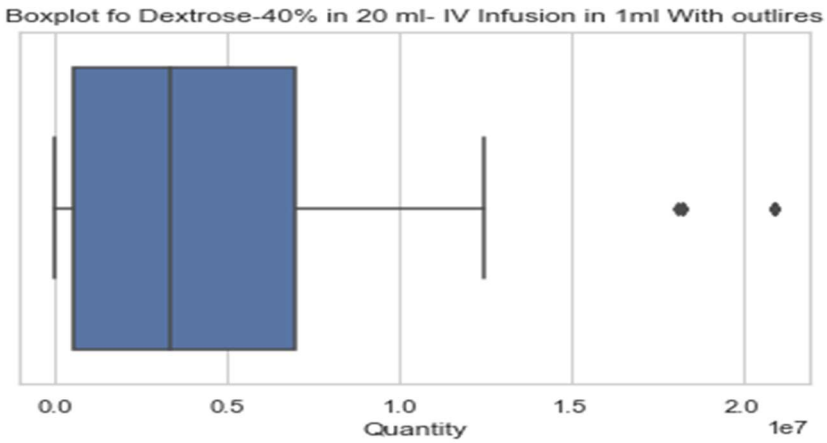



**Figure 3. 10. Boxplot for Dextrose - 40% in 20ml - Iv Infusion**

During the study as on the boxplot above indicated two outliers are detected and removed and replaced by imputation techniques according to table 3.8 shows

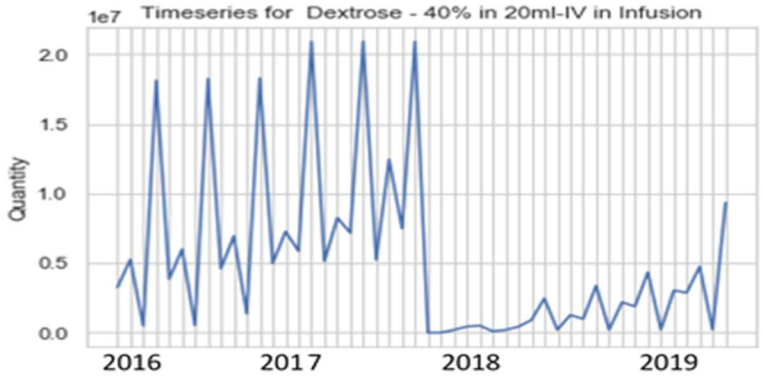



**Figure 3. 11. Time Series for Dextrose - 40% in 20ml - Iv Infusion**

Time Series is a sequence of data-points measured at a regular time-intervals over a period of time. Irregular data does not form Timeseries. It uses statistical methods to analyze time series data and extract meaningful insights about the data. The data points are collected over a period. These data points (past values) are analyzed to forecast a future. Obviously, it is time dependent. Timeseries analysis helps us to recognize the major components in a time series data. From the graph generated by the plotted points, we can see any trends in the data. A trend is a change that occurs in general direction. Having this, the above figure 3.5,3.7,3.9 and 3.11 at some point on the time series graph we can observe that the amount of consumption is 0 means, the item is not stock out at all but it indicates that the stock is not consumed during that time and also 0 indicates annual inventory time, but the stock is available in the stock because these items are mandatory 24 hours of a day, 7 days of a week, and or 12 months of a year.

The models were coded and implemented in the python environment and simulation results were then obtained. As discussed above, boxplot is able to trace the outliers/ variation of the quantity and as we observe on the boxplot graph on different years and months the quantity of the item is become high that means at that month the consumption of the product is high. Therefore, it is suitable to deal with the outliers on timeseries data. The experiment obtained on dataset outliers removed and replaced using imputation.

### 3.4.3. Missing Value Replacement

Missing values refer to the values for one or more attributes in a data was absent or jumped during data collection. This presence of missing values can affect the performance of a prediction constructed using that dataset as a training sample. Scientifically rates of less than 1% missing data are generally considered minor and 1-5% manageable. However, 5-15% needs to refined strategies and more than 15% may cruelly impact any kind of interpretation [36].

Here in this study, the researcher tries to discuss different methods and techniques to handle missing data and the best methods were selected. The following methods were some of them in case of handling missing values:

**Ignore the tuple:** When the class label is missing this method is advisable and the method is not very effective unless the tuple contains several attributes with missing values. It is especially poor when the percentage of missing values per attribute varies considerably [37].

**Use a universal constant to fill in the missing value:** Replace all missing attribute values by the same constant such as a label "Unknown [37].

**Random replace:** replace the missing values with a randomly picked value from the respective column. This technique would be appropriate where the missing values row count is insignificant [38].

**Imputation:** This technique preserved all cases by replacing missing data with a probable value based on other available information. A simple procedure for imputation was to replace the missing value with the mean or median.

For this study, the researcher chooses imputation approaches was used on the mice package. RStudio for imputation (Multivariate Imputation by Chained Equations) for handling the missing values. The reason why the researcher selects this methods was which were a mice package gives functions that can impute continuous, binary, and ordered/unordered categorical data, and imputing each incomplete variable with a separate model that, means in one different values were filled in feature column missed space rather than fill the same values in all missed space. Creating multiple imputations as compared to a single imputation (such as mean) takes care of uncertainty in missing values. Therefore, the MICE imputed data on a variable by variable based on specifying an imputation model per variable [39].

**Fill in the missing value manually:** This approach was time-consuming and may not be feasible given a large dataset with many missing values [37].

**Replace with a summary:** The most commonly used imputation technique. Summarization here is the mean, mode, or median for a respective column [38].

After removed outliers from the crude data. From 196 observation 0.12% missing values was found after removed the IQR as outliers and Imputation techniques are applied to replace the missing values. IQR mostly used technique during outliers removed.

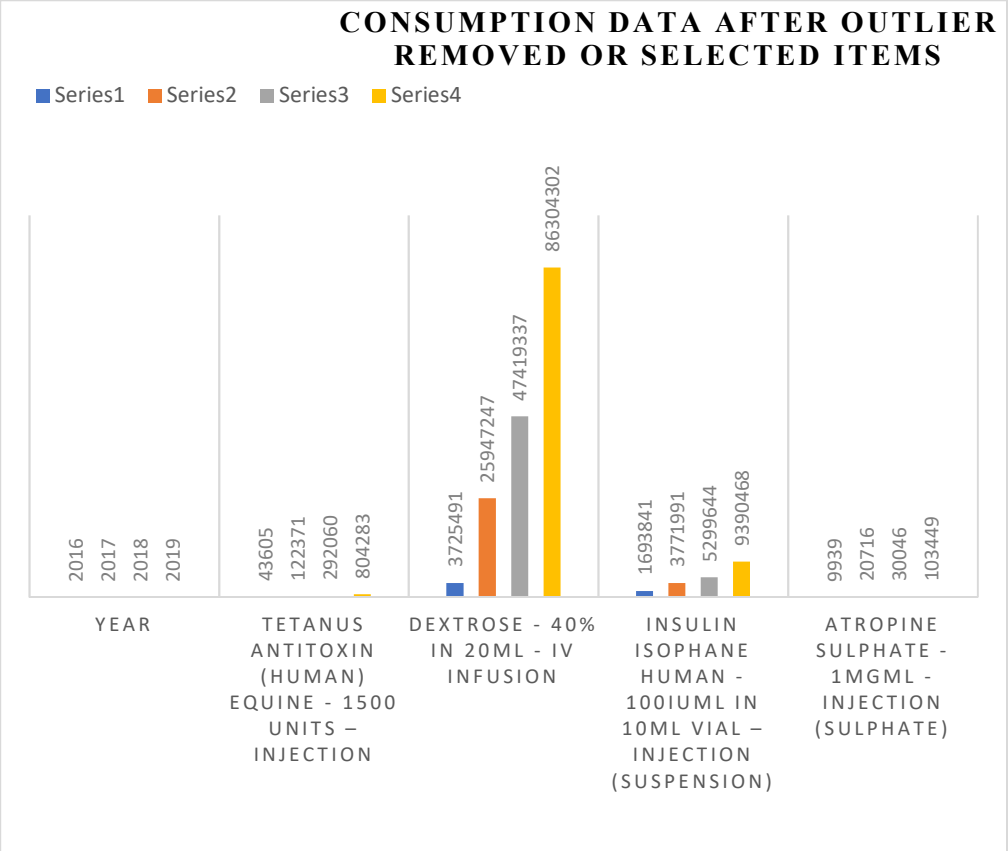



**Figure 3. 12. Annual Consumption of Selected Drugs After Outlier Removed**

The above figure illustrates the annual consumption quantity data per year after replacing the missing values.

## 3.5. Modeling

Demand forecasting in management of thousands of items in a central warehouse is become difficult task. In order to simplify such difficult tasks, it very essential to use the best forecasting method. The researcher used a scientific experimental method.

The study constructs a dataset which contains necessary information to forecast demand of warehouse. The dataset is constructed from real consumption data of a national health commodity from EPSA HCMIS database. Consumption data of 2016, 2017, 2018 and 2019 are used in this research.

Data is cleaned and prepared for further operations. In this research three algorithms such as LSTM, MLP and Auto ARIMA are compared. To assess the

predictability performance of the above three algorithms, four years of data in four different products are applied to all of the algorithms separately. The best algorithm is selected based on Mean Square Error (MSE) value. After comparison the best performed algorithm were selected. In this study, Auto ARIMA, LSTM and MLP algorithms are selected for comparison.

The reason why these three algorithms are selected is, in Auto ARIMA [28] performed better than others algorithms, in [21] LSTM performed better than others algorithm that are compared in that study and in [34] MLP performed better than the rest of other compared with in that study. Therefore, the researcher needed to compare and construct the above three algorithm themselves. The next chapter can be the way how to implement.

### 3.5.1. LSTM

This section describes learning method of our model. We employed RNN using LSTM for prediction. The study decided to use LSTM model cause our data is time series and LSTM is very good at holding long term memories. The prediction in a sequence of test samples can be influenced by an input by an input that was given many time steps before. This network can store or release memory in the go through the gating mechanism.

A LSTM network is a kind of recurrent neural network. A recurrent neural network is a neural network that attempts to model time or sequence dependent behavior such as language, stock prices, electricity demand and so on. This is performed by feeding back the output of a neural network layer at time *t* to the input of the same network layer at time *t + 1*. LSTM step by step implementation.

## Define Network

Neural networks are defined in Keras as a sequence of layers. The container for these layers is the Sequential class. The first step is to create an instance of the Sequential class. Then create the layers and add them in the order that they should be connected. The LSTM recurrent layer comprised of memory units is called LSTM(). A fully connected layer that often follows LSTM layers and is used for outputting a prediction is called Dense().

The first layer in the network must define the number of inputs to expect. Input must be three-dimensional, comprised of samples, timesteps, and features.

- **Samples**. These are the rows in the data.
- **Timesteps**. These are the past observations for a feature, such as lag variables.
- **Features**. These are columns in your data.

The data is loaded as a NumPy array, then convert a 2D dataset to a 3D dataset using the reshape() function in NumPy.

## After defined the network, then compile the network.

Compilation is an efficiency step. It transforms the simple sequence of layers that defined into a highly efficient series of matrix transforms in a format intended to be executed on CPU, depending on how Keras is configured.

Compilation is precompute step for the network. It is always required after defining a model. Compilation requires a number of parameters to be specified, specifically tailored to training your network. Specifically, the optimization algorithm to use to train the network and the loss function used to evaluate the network that is minimized by the optimization algorithm. Keras also supports a suite of other state-of-the-art optimization algorithms that work well with little or no configuration. In this study the adam optimization algorithm was used. ADAM or adam requires the tuning of learning rate.

## Once the network is compiled, it can be fit, which means adapt the weights on a training dataset.

Fitting the network requires the training data to be specified, both a matrix of input patterns, X, and an array of matching output patterns, y. The network is trained using the backpropagation algorithm and optimized according to the adam optimization algorithm and MSE loss function were specified when compiling the model. The backpropagation algorithm requires that the network be trained for a

specified number of epochs or exposures to the training dataset. Each epoch can be partitioned into groups of input-output pattern pairs called batches. This defines the number of patterns that the network is exposed to before the weights are updated within an epoch. It is also an efficiency optimization, ensuring that not too many input patterns are loaded into memory at a time. Once fit, a history object is returned that provides a summary of the performance of the model during training. This includes both the loss and any additional metrics specified when compiling the model, recorded each epoch. Training can take a long time, from seconds to hours to days depending on the size of the network and the size of the training data.

## Once the network is trained, it can be evaluated.

The network can be evaluated on the training data, but this will not provide a useful indication of the performance of the network as a predictive model, as it has seen all of this data before. This study evaluates the performance of the network on a separate dataset, unseen during testing. This will provide an estimate of the performance of the network at making predictions for unseen data in the future. The model evaluates the loss across all of the test patterns, as well as any other metrics specified when the model was compiled, A list of evaluation metrics is returned. As with fitting the network, verbose output is provided to give an idea of the progress of evaluating the model. We can turn this off by setting the verbose argument to 0.

## After satisfied the performance of the fit model, ready to make predictions on new data.

This is as easy as calling the predict() function on the model with an array of new input patterns. The predictions will be returned in the format provided by the output layer of the network. In the case of a regression problem, these predictions may be in the format of the problem directly, provided by a linear activation function.

## Summary

1. **Define Network**: the study constructs an LSTM neural network with a 1 input timestep and 1 input feature in the visible layer, 4 memory units in the LSTM hidden layer, and 1 neuron in the fully connected output layer with a linear (default) activation function.
2. **Compile Network**: the study uses the efficient ADAM optimization algorithm with default configuration and the mean squared error loss function because it is a regression problem.
3. **Fit Network**: the study fit the network for 1,000 epochs and use a batch size equal to the number of patterns in the training set. We will also turn off all verbose output.
4. **Evaluate Network**: the study evaluates the network on the training dataset. Typically, the study evaluates the model on a test or validation set.
5. **Make Predictions**. The study makes predictions for the training input data. Again, typically we would make predictions on data where we do not know the right answer.

### 3.5.2. MLP

Time series regression problems are usually quite difficult, and there are many different techniques. In this study we use time series regression using MLP neural network. The method creates training data and the training data is also re-normalized by using min-max score. The first four values on each line are used as predictors. The fifth value is the quantity count to predict. In other words, each set of four consecutive quantity counts is used to predict the next count. The size of the rolling window used here, 4, was determined by trial and error. The method creates MLP neural network with four input nodes, 12 hidden processing nodes, and a single output node.

There's just one output node because time series regression predicts just one-time unit ahead. The number of hidden nodes in a neural network must be determined by trial and error. The MLP neural network has (4 * 12) + (12 * 1) = 60 node-to-

node weights and (12 + 1) = 13 bias which essentially define the MLP neural network model. Using the rolling window data, the method trains the network using the basic stochastic back-propagation algorithm with a learning rate set to 0.01 and a fixed number of iterations set to 10,000. Both the learning rate and number of iterations are free parameters and their values must be determined by experimentation.

The training algorithm uses back-propagation, which is a form of stochastic gradient descent, with a batch size of one. The error function used is Mean Square Error because the predicted output value and known correct output value from the training data are numeric. After training is completed, the method calculates and displays a few actual quantities counts and quantities counts predicted by the neural model.

### 3.5.3. Auto ARIMA

ARIMA is a very powerful model for forecasting time series data, the data preparation and parameter tuning processes end up being really time consuming. Before implementing ARIMA, you need to make the series stationary, and determine the values of p and q using the plots we discussed above. Auto ARIMA makes this task really simple Below are the steps we should follow for implementing auto ARIMA:

1. Load the data: This step will be the same. Load the data into your notebook
2. Preprocessing data: The input should be univariate, hence drop the other columns
3. Fit Auto ARIMA: Fit the model on the univariate series
4. Predict values on validation set: Make predictions on the validation set
5. Calculate MSE: Check the performance of the model using the predicted values against the actual values

We completely bypassed the selection of p and q feature as we can see. we implement Auto ARIMA using a consumption dataset.

### 3.5.4. Data Split

Data splitting includes dividing the data into an explicit training dataset used to load the data into the model and an unseen test dataset used to evaluate the model's performance on unseen data. The data was split to train dataset and test dataset in python. The dataset was randomly categorized into two groups, training, and test groups for model validation as recommended [40]. In this study, Data are divided in to two group randomly.

### 3.5.5. Mean Square Error

On the study the performance and errors of the algorithm s measured by MSE. The Mean Square Error or MSE is applied measure of the differences between numbers population values and samples which is predicted by estimator. The MSE describes the sample standard deviation of the differences between the predicted and observed values. The differences are known as residuals. During calculation done over the data sample that was used to estimate and known as prediction errors when calculated out of sample. The MSE aggregates the magnitudes of the errors in predicting different time into a single measure of predictive power [41]

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(x_t - \hat{x}_{t-1,t})^2 \quad \text{..... (1)}$$

## 3.6. Tools

using Jupiter for prediction and also the study was used Python 3.6 to build the model on Jupiter platform, and MS Excel to convert the data in to comma delimited (CSV) format, and lastly the data was loaded to the testing algorithms on python.

There are various techniques followed in order to achieve a given algorithms goal. Since this is a study which attempts to build a model that predicts the demand prediction, there is a need of discussing the techniques used in prediction and building a model. For this reason, the detailed discussion of the techniques is followed in each sub section.

### 3.6.1. Anaconda

Anaconda is a tool we are used during our experimentation its open source data science package with a community of over 6 million users it is easy to download and install, and it is supported on Linux, MacOS, and Windows [27].

Anaconda is free Python distribution, including over 195 of the most popular Python packages for science, math, and data analysis. Anaconda includes Python 2.7/Python 3.6 and cross-platform Python packages, as well as tools for integration with Excel.

#### 3.6.1.1.Python

Python starts since1991, Adoption of Python for scientific computing in both industry applications and academic research has increased significantly since the early 2000s. For data analysis and interactive, exploratory computing and data visualization. In recent years, Python's improved library support (primarily pandas) has made it a strong alternative for data manipulation tasks [42]. Combined with Python's strength in general purpose programming, it is an excellent choice as a single language for building data-centric applications. The scientist's needs python to Get data (simulation, experiment control), Manipulate and process data, visualize results, quickly to understand, but also with high quality figures, for reports or publications. Python Rich collection of already existing bricks of classic numerical methods, plotting or data processing tools. We don't want to re-program the plotting of a curve, a Fourier transformer a fitting algorithm. Python is Efficient code Python numerical modules are computationally efficient. But needless to say, that a very fast code becomes useless if too much time is spent writing it. Python aims for quick development times and quick execution times. Python is a universal language used for many different problems. Learning Python avoids learning a new software for each new problem. Python core numeric libraries NumPy, SciPy, Pandas and Matplotlib. Specific packages of python are Mayavi, pandas, stats models, seaborn, sympy, scikit-image and scikit-learn [43].

#### 3.6.1.2. Jupyter

The Jupyter notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text.

### Numpy

NumPy, short for Numerical Python, has long been a cornerstone of numerical computing in Python. It provides the data structures, algorithms, and library glue needed for most scientific applications involving numerical data in Python.
Beyond the fast array-processing capabilities that NumPy adds to Python, one of its primary uses in data analysis is as a container for data to be passed between algorithms and libraries. For numerical data, NumPy arrays are more efficient for storing and manipulating data than the other built-in Python data structures. Also, libraries written in a lower-level language, such as C or Fortran, can operate on the data stored in a NumPy array without copying data into some other memory representation. Thus, many numerical computing tools for Python either assume NumPy arrays as a primary data structure or else target seamless interoperability with NumPy.

### Matplotlib

Matplotlib is the most popular Python library for producing plots and other two-dimensional data visualizations. It was originally created by John D. Hunter and is now maintained by a large team of developers. It is designed for creating plots suitable for publication. While there are other visualization libraries available to Python programmers, Matplotlib is the most widely used and as such has generally good integration with the rest of the ecosystem. In our study we use it for visualization purpose.

### Pandas

The pandas name itself is derived from panel data, an econometrics term for multidimensional structured datasets, and a play on the phrase Python data analysis itself

Pandas provides high-level data structures and functions designed to make working with structured or tabular data fast, easy, and expressive. Since its emergence in 2010, it has helped enable Python to be a powerful and productive data analysis environment.

Pandas for data manipulation, preparation, and cleaning is such an important skill in data analysis.

### SciPy

**SciPy** stands for Scientific Python. SciPy is built on NumPy. Collection of packages addressing a number of different standard problem domains in scientific computing. It is one of the most useful libraries for variety of high-level science and engineering modules like discrete Fourier transform, Linear Algebra, Optimization and Sparse matrices.

### Scikit-learn

For machine learning. Built on NumPy, SciPy and Matplotlib, this library contains a lot of efficient tools for machine learning and statistical modeling including linear regression, SVM, nearest neighbors, feature selection, matrix factorization, Grid search, cross-validation, metrics, Feature extraction, and normalization.

### Stats models

Stats models is a statistical analysis package that was seeded by. Compared with scikit-learn, Stats models contains algorithms for classical (primarily frequentist) statistics and econometrics. This includes and an in our study used submodules are:

Regression models: Linear regression, generalized linear models, robust linear models, linear mixed effects models, etc. and Time series analysis: AR, ARMA,

ARIMA, VAR, and other models and to Visualization of statistical model results Stats models is more focused on statistical inference, providing uncertainty estimates and *p*-values for parameters. scikit-learn, by contrast, is more prediction-focused [44].

# CHAPTER FOUR: EXPERIMENT, RESULT AND DISCUSSION

This section describes the results of what we proposed in chapter three and the results were obtained by applying the three algorithms, MSE. First step is to collect the data. Second step is to inspect the redundancy and outliers using python. Finally, input variable and output variable are selected based on literature [21].

In this study we have considered consumption dataset from EPSA central Agency. To show what the actual dataset looks like and to made the dataset essay to understand we applied a data visualization technique. This phase also displays to validate the use of the proposed algorithms when estimating the demand of different health commodity.

The study is performed on health commodity consumption with 4 years of data starting from 2016 to 2019. Three forecasting algorithms was compared on this study, ARIMA, LSTM and Multi-Layer Perceptron methods was applied to four selected vital drugs, both methods considered the four vital drugs for forecasting. MSE of the four selected drugs was compared in three algorithms.

Then, this was explored which algorithm is the best in terms of MSE result and last, the forecast accuracy result for each algorithm was analyzed. The error measure used are MSE. The study was also investigated whether the proposed algorithms performs superior to each other when forecasting demand performed for selected all drugs. The following graphs and tables illustrate the forecast accuracy of the three algorithms in the next year annual demand for the selected drug.

## 4.1. Modeling

### 4.1.1. Long-Short Term Memory LSTM

Figure 4.1. The run where the LSTM predict for tetanus Antitoxin (Human) Equine - 1500 Units – Injection forecast MSE.

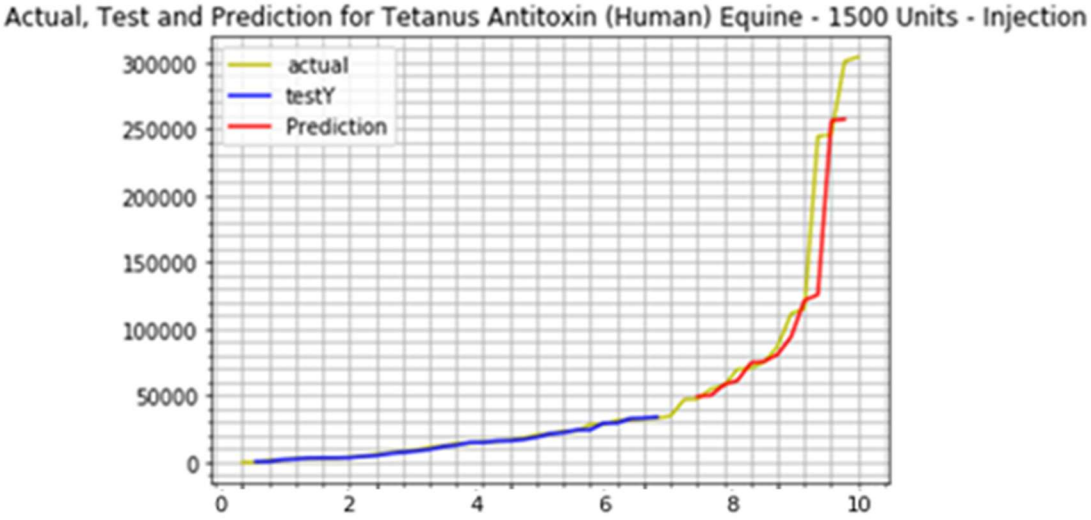



**Figure 4. 1. Predicted Values for Tetanus Antitoxin (Human) Equine - 1500 Units – Injection using LSTM.**

As we can see on the above graph the line yellow shows the actual data set, the blue one shows the training data set and the read one shows the unseen test dataset. The line plot is showing the observed values compared to the rolling forecast predictions overall our forecasts align with the true values as well. The run where the LSTM predict for tetanus Antitoxin (Human) Equine - 1500 Units – Injection, taking the square root of the performance scores we can see the MSE on the training dataset was 1,124.55 quantity (in thousands per year) and the average error on the unseen test set was 37,278.99quantity (in thousands per year).

| Item | Algorithm | MSE Train lose | MSE Test lose | Accuracy |
|---|---|---|---|---|
| Tetanus Antitoxin (Human) Equine - 1500 Units – Injection | LSTM | 1124.55 | 37278.99 | 80% |
| Dextrose - 40% in 20ml - Intravenous Infusion | | 220839.49 | 1673443.63 | 90% |
| Insulin Isophane Human - 100IUml in 10ml Vial – Injection (Suspension) | | 48223.08 | 65714.02 | 90% |
| Atropine Sulphate - 1mgml - Injection (Sulphate) | | 137.57 | 627.37 | 93% |

**Table 4. 1. MSE for LSTM Algorithm**

### 4.1.2. Multi-Layer Perceptron

Figure 4.2. The run where the Multi-Layer Perceptron predict for tetanus Antitoxin (Human) Equine - 1500 Units – Injection and forecast MSE.

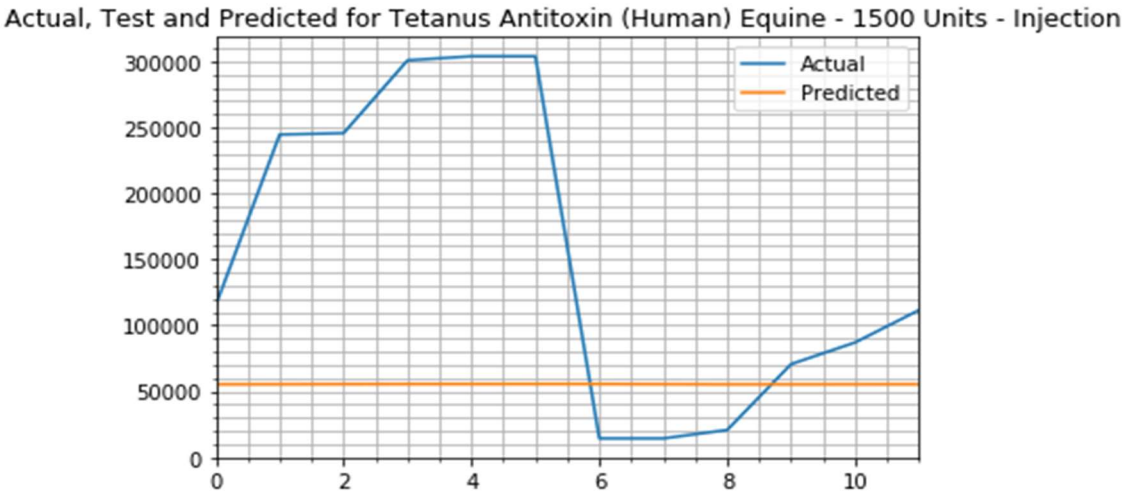



**Figure 4. 2. Predicted Values for Tetanus Antitoxin (Human) Equine - 1500 Units – Injection using Multi- Layer Perceptron.**

As we can see on the above graph the line red shows the actual data set, the blue one shows the unseen test dataset. The line plot is showing the observed values compared to the rolling forecast predictions overall our forecasts align with the

actual values very bad. The run where the Multi-Layer Perceptron predict for tetanus Antitoxin (Human) Equine - 1500 Units – Injection, taking the square root of the performance scores we can see the MSE on the training dataset was 6,978.78 quantity (in thousands per year) and the average error on the unseen test set was 6,954.16 quantity (in thousands per year).

| Item | Algorithm | MSE Train Score | MSE Test Score | Accuracy |
|---|---|---|---|---|
| Tetanus Antitoxin (Human) Equine - 1500 Units – Injection | Multi-Layer Perceptron | 6,978.78 | 6,954.16 | 89.1% |
| Dextrose - 40% in 20ml - Intravenous Infusion | | 391,075.97 | 389,354.57 | 93.5% |
| Insulin Isophane Human - 100IUml in 10ml Vial – Injection (Suspension) | | 89,145.00 | 60,312.77 | 63.0% |
| Atropine Sulphate - 1mgml - Injection (Sulphate) | | 1227.81 | 866.10 | 56.5% |

**Table 4. 2. MSE Score for Multi-Layer Perceptron Algorithm**

### 4.1.3. Forecasting by ARIMA for Selected Item

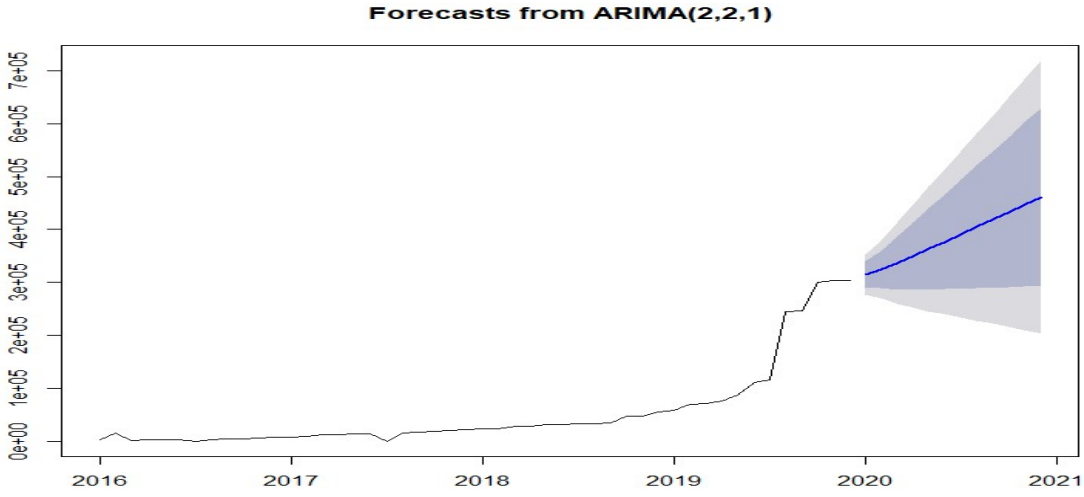


**Figure 4. 2 Predicted Values for Tetanus Antitoxin (Human) Equine - 1500 Units – Injection using Auto ARIMA.**

The above graph shows the forecasting of the future one-year consumption of pharmaceuticals where the shaded is the confidence of the forecasting which is shown below in the table with forecasted values.

| Month | Tetanus Antitoxin (Human) Equine - 1500 Units – Injection | | Insulin Isophane Human – 100IUml in 10ml Vial-Injection (Suspension) | | Dextrose-40% in 20ml-Intravenous infusion | | Atropine Sulphate-1mg/ml | |
|---|---|---|---|---|---|---|---|---|
| | Lo | Hi | Lo | Hi | Lo | Hi | Lo | Hi |
| Jan-20 | 315043 | 353295 | 1248614 | 1362903 | 21813637 | 23637969 | 13201 | 14583 |
| Feb-20 | 324043 | 378182 | 1290733 | 1460039 | 22635326 | 26218052 | 13687 | 15741 |
| Mar-20 | 337419 | 415331 | 1332852 | 1549723 | 23457015 | 28330783 | 14173 | 16812 |
| Apr-20 | 349647 | 446656 | 1374971 | 1636492 | 24278704 | 30293561 | 14659 | 17850 |
| May-20 | 363630 | 481858 | 1417090 | 1722002 | 25100393 | 32186143 | 15145 | 18875 |
| Jun-20 | 377032 | 514762 | 1459209 | 1807047 | 25922082 | 34042361 | 15631 | 19896 |
| Jul-20 | 391149 | 549095 | 1501327 | 1892071 | 26743771 | 35879908 | 16117 | 20919 |
| Aug-20 | 404986 | 582552 | 1543446 | 1977340 | 27565460 | 37709193 | 16602 | 21947 |
| Sep-20 | 419118 | 616655 | 1585565 | 2063023 | 28387149 | 39536819 | 17088 | 22981 |
| Oct-20 | 433119 | 650509 | 1627684 | 2149233 | 29208838 | 41367199 | 17574 | 24022 |
| Nov-20 | 447243 | 684754 | 1669803 | 2236046 | 30030527 | 43203391 | 18060 | 25072 |
| Dec-20 | 461307 | 719037 | 1711922 | 2323514 | 30852216 | 45047577 | 18546 | 26132 |

**Table 4. 3. Forecasting by Auto ARIMA for Selected Item**

## 4.2. Evaluation

In this section, the performance analysis of prediction algorithms of ARIMA, LSTM and Multi-Layer Perceptron were carried out with a common dataset using Python tool on Jupiter and R studio platform. In order to measure the error of algorithms the study use MSE metric. Finally, the best results were selected depending on the result of the metric in table 4.4 the MSE of the three algorithms. As can be seen in this table, ARIMA gives better results than two algorithms four our problem.

| Algorithm | Error and accuracy techniques | Tetanus Antitoxin (Human) Equine - 1500 Units – Injection | Dextrose - 40% in 20ml - Intravenous Infusion | Insulin Isophane Human - 100IUml in 10ml Vial – Injection (Suspension) | Atropine Sulphate - 1mgml - Injection (Sulphate) |
|---|---|---|---|---|---|
| **Gradient Descent** | **MSE Train loss** | 6,978.78 | 391,075.97 | 89,145.00 | 1,227.81 |
| | **MSE Test loss** | 6,954.16 | 389,354.57 | 60,312.77 | 866.1 |
| **LSTM** | **Train loss** | 1,124.55 | 220,839.49 | 48,223.08 | 615.05 |
| | **Test loss** | 37,278.99 | 1,673,443.63 | 65,714.02 | 1,771.65 |
| **ARIMA** | **MSE loss** | 3,747 | 173,979 | 36,290 | 329 |

**Table 4. 4. MSE loss for Algorithms**

## 4.3. Discussion

The dataset used in this study has many missing values and outliers the features in the dataset 13 features were counted. Importance features selection was very difficult that feed to prediction models from 13 features. Therefore, the dataset was preprocessed by using data preprocessing techniques, depending on literature review and expert recommendation 5 features were selected. After 5 features selected using literature review two features were feed to algorithms. In case of filling or handling missing values, most researchers used mean values of the feature itself. That means for the missed value in the features the same value was filled and this may be not all records in the features not the same. Here in this study, we used mice package in RStudio. for multiple imputations that was given functions that can impute continuous, binary /unordered categorical data, and imputing each incomplete variable with a separate model.

Accuracy of demand forecasting can be enhanced through efficient and effective data collection, analysis and prediction through using best mode of prediction models and techniques, forecasting the demand and perform prediction experimentally to evaluate the performance of two supervised machine learning algorithms and ARIMA was used as the testing design for this study. First, the LSTM and Multi-Layer Perceptron algorithm have been trained on a set of training

sample with the preprocessed dataset. On ARIMA and two algorithm prediction was performed as well.

According to the above tables table 4.3, the MSE for Atropine Sulphate - 1mgml - Injection (Sulphate) is 329, Tetanus Antitoxin (Human) Equine - 1500 Units – Injection is 3,747, Insulin Isophane Human - 100IUml in 10ml Vial – Injection (Suspension) is 36,290 and Dextrose - 40% in 20ml - Intravenous Infusion is 173,979 , when we consider the demand prediction for all drugs the accuracy final result is highly optimized and the level of wastages, expiry and stockout is significantly reduced, the availability can be improved thus using ARIMA method is a good for forecasting. The result showed the prediction accuracy of ARIMA was best. To make good the accuracy of the two algorithm we perform different tasks starting from changing the number of iteration and sampling percentage.

RQ1. What methods and techniques had been used at EPSA to forecast the future demand?

Having the research question EPSA has techniques to forecast the future demands, however the study indicates that the forecasting accuracy of the agency is poor.

RQ2. What are the major limitations in the current EPSA to give an accurate forecasting for future demand?

According to the study the agency uses simple statistical forecasting method using excel spread sheet rather than using modern machine learning methods. The agency uses the consumption dataset that was collected manually to forecast the future demand. This indicates that during data collection there is chance of poor data quality and false data this leads the agency forecast accuracy poor.

RQ3. Which ML algorithms more appropriate to forecast the future demand of EPSA?

According to the study result, forecasting accuracy of ARIMA was best than MLP and LSTM.

# CHAPTER FIVE: CONCLUSIONS, RECOMMENDATION AND FUTURE WORK

## 5.1. Conclusions

On this study prediction technique was used to forecast demand health commodity. The study compares three algorithms. According to the analysis result of the prediction experiments on selected item ARIMA performs with best MSE result according to table 4.3 than LSTM and Gradient Descent. We observe that according to experimental result ARIMA is most suitable method for this type of dataset to make real the generated result from prediction.

Conclusion form the aforementioned findings; the study concludes that cleaning the dataset carefully before applying the dataset, time series plot and boxplot also one of the most important way to visualize the dataset and to examine or investigate over all the dataset and normalizing process is important to make dataset suitable for the prediction these two findings are need very high attention before prediction.

This study reveals that the candidate sectors have no full of knowledge on demand forecasting and they only use excel sheet to forecast the demand.

The final conclusion is that, the EPSA have to be give attention on its demand prediction method and data utilization. In addition to that the results from the study have shown that, the problem of forecasting leads the agency high wastage rate and capital crises as national level.

The study is done for academic purpose but it can be implemented for the Agency, branches health facility and pharmaceutical commodity manufacturers.

The data used for the study is five-year data EPSA central office; therefore, if EPSA partners and researches are conducted on this area and on EPSA data helps for different health sector related study it can be helpful for the agency to implement the application of machine learning.

In addition, there are different algorithms are available that can help to perform demand prediction. So other better algorithms can be selected by performing better than these algorithms used in this study to predict demand. Therefore, it is a good

effort if anyone can apply and get best algorithm to build better prediction model for this type of data.

Finally, the following conclusions are drawn from the finding with regard to the research questions.

A supervised algorithms and statistical method have been executed to gain a good and acceptable forecast accuracy.

- A method for successfully forecast demand using machine learning techniques has been made possible.
- An algorithm that enabled demand forecasting algorithms are comparing each other to fined best one.

Besides, the selected method has also enabled forecasting the demand accurately under time series dataset.

## 5.2. Recommendation

Based on the research study conducted, in order to get the advantage of accuracy in demand forecasting the researcher recommends the following:

1- Improve its accuracy in demand forecasting by deploy ARIMA method for demand forecasting and reduce stockouts and insure availability.

2- Imputation technique in replacing an average monthly consumption-based stock out management is useful.

3- Annual inventory period for efficient inventory management system through using demand forecasting method.

4- Mechanism that handle seasonal and event-based demand fluctuation management system.

5- EPSA should deploy a web-based data tracking and processing mechanism.

This study contributes for EPSA a new scientific method forecasting health commodity demand.

## 5.3. Future work

In this section we insight a key point that remain a challenge in pharmaceuticals demand forecasting, of course the challenge in this work. The data were collected from HCMIS data base only. A multi factor to forecast pharmaceutical demand will bring a significant role in pharmaceuticals demand forecasting. Based on that, the researcher believes future study should consider it to achieve better result in including factors such as demographic data, seasonality and bullwhip effect pharmaceuticals demand forecasting.

## REFERENCES


[1] F. v. Sommeren, "Improving forecast accuracy," December 5, 2011.

[2] R. Lug, "A study to examine the importance of forecast Accuracy to supply chain performance, the contributing factors and the improvement enablers in practice," 2015.

[3] A. K. a. S. S. I. J. M. E. &. Rob, "Demand forecasting using neural network for supply chain management," 2015.

[4] B. A. A. B. Paul W. Murray, "Forecasting Supply Chain Demand by Clustering Customers," Canada, 2015.

[5] H. R. Y. F. T. Gökçe Candan, "Demand Forecasting In Pharmaceutical Industry Using Artificial Intelligence: Neuro-Fuzzy Approach," March 2014.

[6] A. S. A. L. N. G. a. U. S. Aditya Chawla, "Demand Forecasting Using Artificial Neural Networks—A Case Study of American Retail Corporation," 2019.

[7] S. A. A. S. S. Lakshmi Anusha, "Demand Forecasting for the Indian Pharmaceutical Retail: A Case Study," *Journal of Supply Chain Management Systems ,* vol. Volume 3 , no. Issue 2, April 2014.

[8] U. |. D. PROJECT, "Ethiopia: National Survey of the Integrated Pharmaceutical Logistics System," U.S. Agency, FEBRUARY 2015.

[9] M. lera, "Standard Operating Procedures (SOP) Manual forthe Integrated Pharmaceuticals Logistics System in Health Facilities of Ethiopia Second Edition," PFSA, Addis Ababa, Ethiopia, November, 2015.

[10] R. H. a. M. Sultan, "Impact of Poor Forecasting Accuracy," 2014.

[11] M. S. N. A. a. A. S. Petter Næss, "Forecasting inaccuracies: A result of unexpected events, optimism bias, technical problems, or strategic misrepresentation?," *The Journal of Transport and Land Use,* vol. Vol. 8 No. 3, 2015.

[12] U. |. D. PROJECT, "Quantification of Health Commodities A Guide to Forecasting and Supply Planning for Procurement," JUNE 2014.

[13] A. Guazzelli, "Predicting the future, Part 2: Predictive modelingtechniques," 2012.

[14] S. S.-S. a. S. Ben-David, UNDERSTANDING MACHINE LEARNING From Theory to Algorithms, New York: Cambridge University Press, 2014.

[15] T. O. Ayodele, "Types of Machine Learning Algorithms," 2010.

[16] S. M. R. R. Sakshi Kohli, "BASICS OF ARTIFICIAL NEURAL NETWORK," 2014.

[17] R. B. S. Agatonovic-Kustrin *, "Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research," 2000.

[18] K. Suzuki, Introduction to the Artificial Neural Networks, InTech, 2011.

[19] D. E. a. J. S. Felix A. Gers, "Applying LSTM to Time Series Predictable through Time-Window Approaches," 2001.

[20] G. Petnehazi, "Recurrent Neural Networks for Time Series Forecasting," 2019.

[21] "Recurrent Neural Networks for Time Series Forecasting," 2019.

[22] "file:///C:/Users/Natking/Desktop/ad%20data/Time%20Series%

20Prediction%20with%20LSTM%20Recurrent%20Neural%20Networks%

20in%20Python%20with%20Keras.html," [Online].

[23] T. P. Kehlog Albran, "Demand Forecasting with Regression Models," 2017.

[24] L. Nayal, "Regression Analysis T o F orecast the Demand of New Single F amily Houses in USA," 2015.

[25] A. V. A. S. Galina Merkuryevaa, "Demand forecasting in pharmaceutical supply chains: A case study," *ScienceDirect,* 2018.

[26] 2. D. F. a. 3. A. F. 1st Angeliki Papana, "Forecasting the consumption and the purchase of a drug," Greece, 2012.

[27] S. M. a. M. M. Rouzbeh Ghousi, "Application of Data Mining Techniques in Drug Consumption Forecasting to Help Pharmaceutical Industry Production Planning," Tehran, 2012.

[28] J. F. Z. A. H. E. M. a. A. L. Jamal Fattah, "Forecasting of demand using ARIMA model," 2018.

[29] D. S. A. G. N. P. K. S. S. K. Srayanta Mukherjee, "Associative and Recurrent Mixture Density Networks for eRetail Demand Forecasting," 2018.

[30] A. S. P. G. Kalyan Mupparaju, "A Comparative Study of Machine Learning Frameworks for Demand Forecasting," 2018.

[31] S. G. O. Irem Islek, "A Decision Support System for Demand Forecasting based on Classifier Ensemble," vol. 13 DOI.

[32] S. M. Sanjita Jaipuria, "An improved demand forecasting method to reduce bullwhip effect in," 2014.

[33] K. Kandananond, "Consumer Product Demand Forecasting based on Artificial Neural Network and Support Vector Machine," vol. Vol:6, 2012.

[34] S. C. a. F. Meneguzzi, "Forecasting Demand with Limited Information Using Gradient Tree Boosting," 2017.

[35] ACAPS, "Data cleaning-dealing with messy data," 2016.

[36] B.Lantz, "Machine Learning with R Second Edition; Discover how to build machine learning algorithms, prepare data, and dig deep into data prediction techniques with R," 2015.

[37] L.Brett, "Machine Learning with R: Learn how to use R to apply pwerful machine learning,," 2013.

[38] S. Manohar, "Mastering Machine Learning with Python in Six Steps: A Practical Implementation Guide to Predictive Data Analytics Using Python," 2017.

[39] A.Tsai, "Methods for Multiple imputing Analysis with R," 1/12/2019.

[40] E. D. a. L. Tommy, "Statistics and Machine Learning in Python," vol. Release 0.1, 2018.

[41] C. D. ROBERT SIWERZ, "Predicting sales in a food store department using machine learning," 2017.

[42] W. McKinney, Python for Data Analysis, United States of America.: O'Reilly Media, 2013.

[43] "Gaël Varoquaux Emmanuelle Gouillart Olaf VahtrasScipy Lecture Notes www.scipy-lectures.org . [Online] 2017," [Online].

[44] G. V.Kass, "Chi-square Automatic Interaction Detector (CHAID) was a technique in," 1980.

[45] b. *. G. N. a. S. K. c. M.-L. N. a. C. P. a. d. F. M. a. A. F. Cyril Voyant a, "Machine learning methods for solar radiation forecasting," 2017.

[46] S. G. O. Irem Islek, "A Decision Support System for Demand Forecasting based on Classifier Ensemble," 2017.

[47] "Gaël Varoquaux Emmanuelle Gouillart Olaf VahtrasScipy Lecture Notes www.scipy-lectures.org," 2017. [Online].

[48] J. Verzani, Getting Started with R Studio, 2011.

[49] A. kowalczyk, Support Vector Machines succinctly, 2501 Aerial center parkway: Syncfusion, Inc., 2017.

[50] T. E. a. M. Pontil, "WORKSHOP ON SUPPORT VECTOR MACHINES: THEORY AND APPLICATIONS," Cambridge, 2001.

[51] Q. S. a. S. A. RO´ NA´ N DALY1, "Learning Bayesian networks: approaches and issues," Vols. 26:2, 99–157, 2011.

[52] W.-H. C. a. J.-Y. Shih, "Comparison of support-vector machines and back propagation neural networks in forecasting the six major Asian stock markets," 2006.

[53] L. C. Francis E.H. Tay ∗, "Application of support vector machines in Financial time series forecasting," 2001.

[54] M. H. S. a. M. S. Hosen, "Prediction on Large Scale Data Using Extreme Gradient Boosting," 2016.

[55] W. McKinney, Python for Data Analysis, United States of America: O'Reilly Media, 2013.

[56] E. Duchesnay and L. Tommy, Statistics and Machine Learning in Python, Release 0.1, 2018.

[57] S. a. K. Groothuis-Oudshoorn, "Mice:Multivariate imputation by chained equations in R," 2011.

[58] B. Jason, "Machine Learning Mastery With R," 2017.

[59] W. M. N. R. a. W. R. R. P. Wiesław, "Application of all-relevant feature selection for the failure analysis of parameter-induced simulation crashes in climate models,," 2016.

[60] T. Brook, A. Suleman, E. Noah, D. Legesse, S. Syed-Abdul and K. Mihiretu, ""Determinants and development of a web-based child mortality prediction model in resource-limited settings: A data mining approach": Ethiopia.," *Computer Methods and Programs in Biomedicine,* vol. 140, p. 45–51, 2017.

[61] D. Bedada, ""Determinant of Under-Five Child Mortality in Ethiopia.," *American Journal of Statistics and Probability,* vol. 6, no. 4, p. 198, 2017.

[62] "https://www.geeksforgeeks.org/gradient-descent-algorithm-and-its-variants/," [Online].

[63] "https://hackernoon.com/gradient-descent-aynk-7cbe95a778da," [Online].

[64] M. K. Jiawei Han and Micheline KamberAcademic Press, " Data Mining: Concepts and Techniques," 2001.

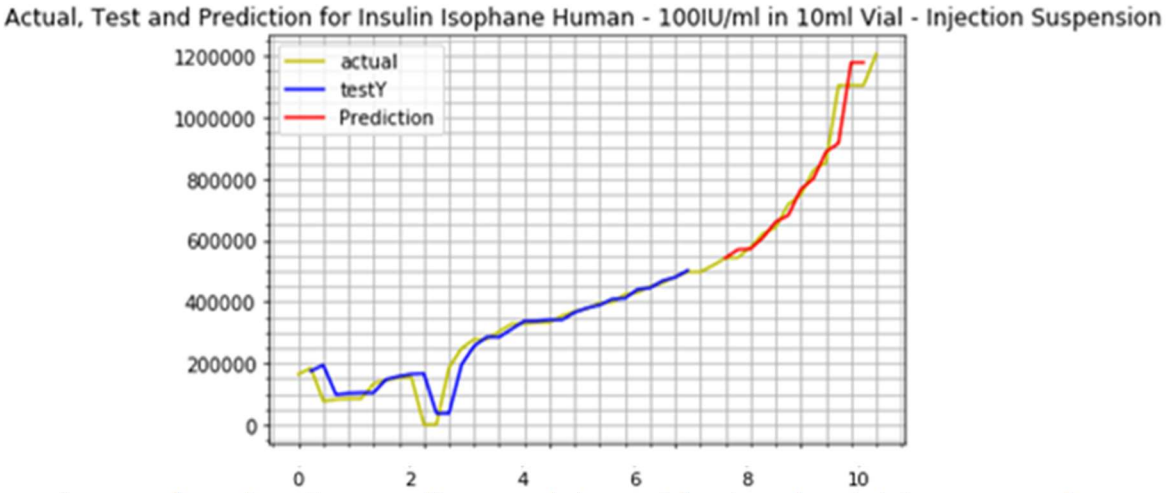
[65] N. L. a. a. Zahavi, "The Data Mining and Knowledge Discovery Handbook," Vols. (pp.1189-1220) , January 2010.

[66] T. e. m. h. a. i. critics, "Malkiel, Burton G," *Journal of Economic Perspectives,* Vols. pp. 107-111, 2003. .

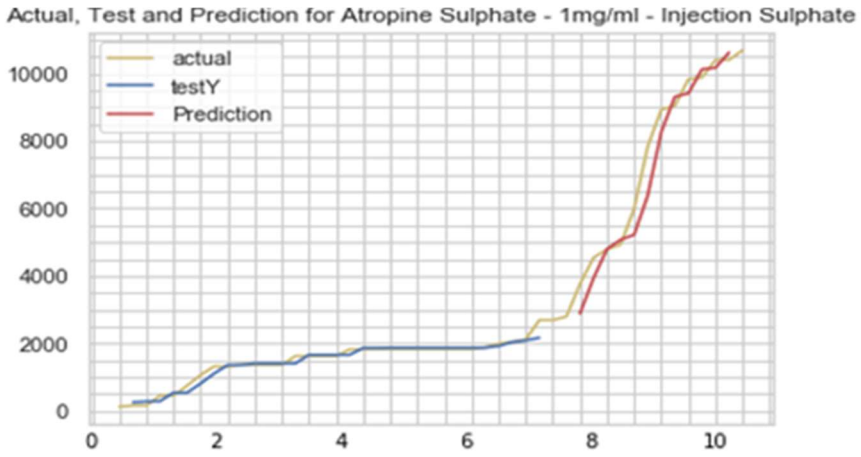
[67] S. Paliyawan, "Market Direction Prediction using Data Mining Classification Conference Paper," December 2014.

# APPENDIX A

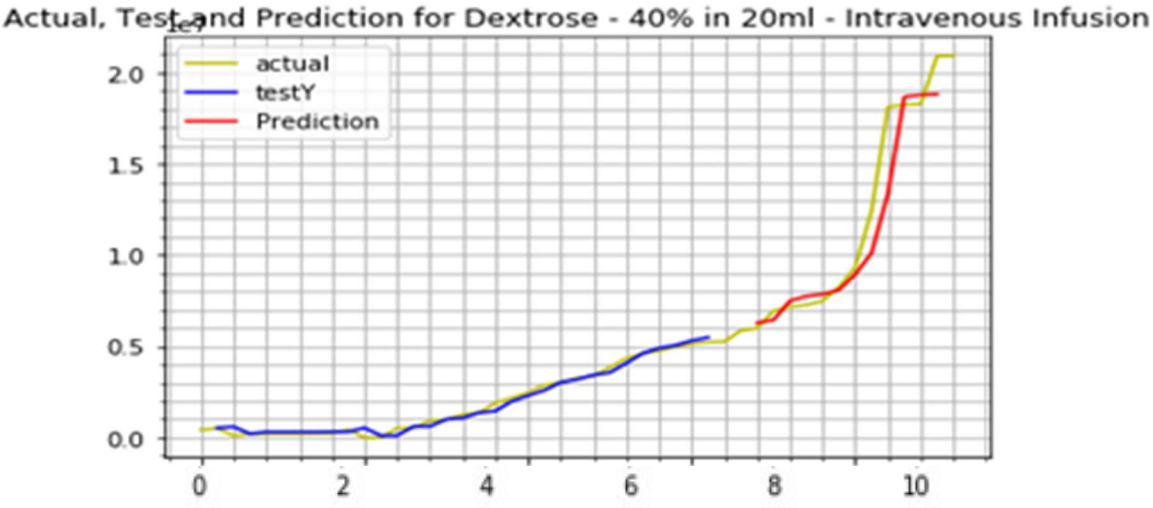
Figure 4.2. The run where the LSTM predict annual demand for Insulin Isophane Human - 100IUml in 10ml Vial – Injection (Suspension) forecast MSE.




The run where the LSTM predict annual demand for Atropine Sulphate - 1mgml - Injection (Sulphate) forecast MSE.

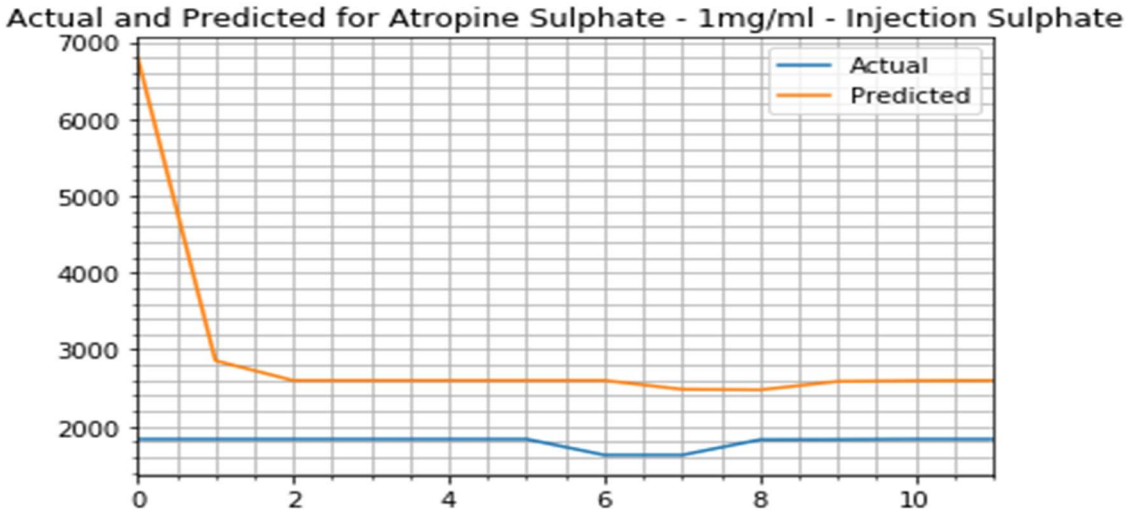


The run where the LSTM predict for Dextrose – 40% in 20ml – Intravenous Infusion forecast MSE.

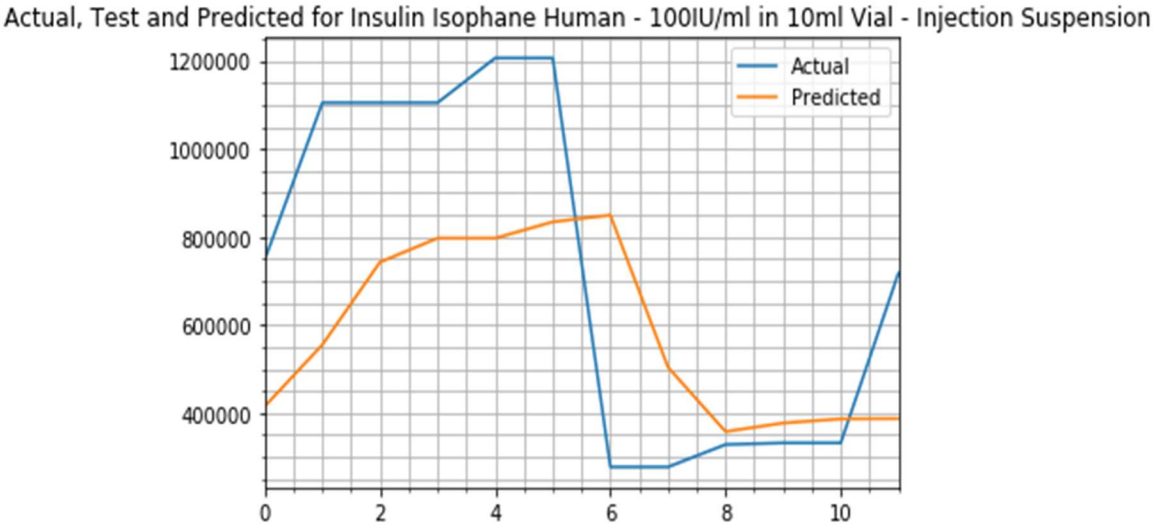


# APPENDIX B

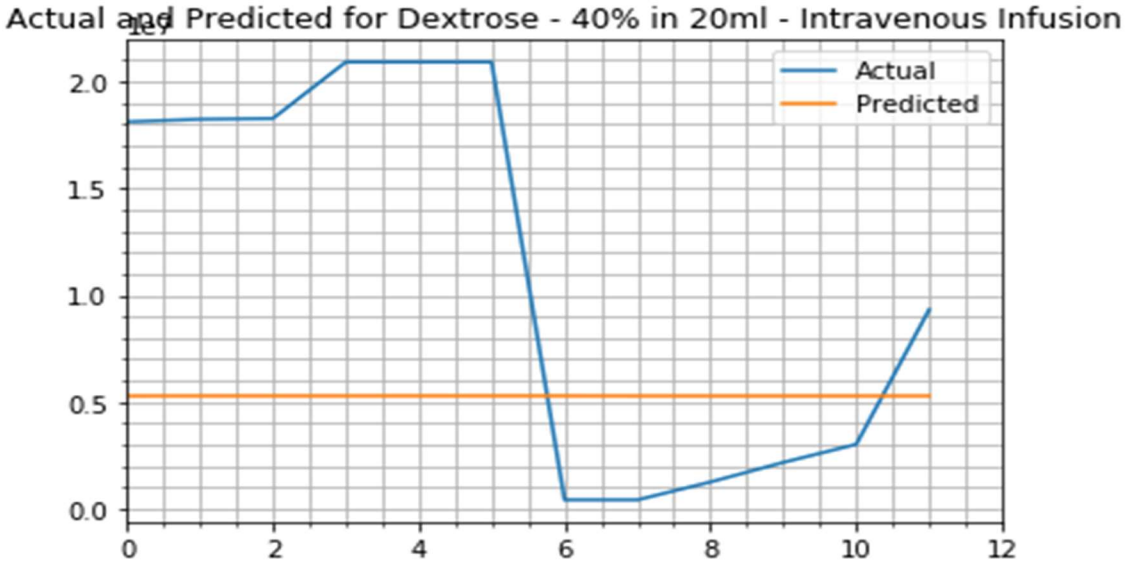
The run where the Multi-Layer Perceptron Predicted for Atropine Sulphate-1mg/ml-Injection Sulphate and forecast of MSE.




The run where the Multi-Layer Perceptron predict for Insulin Isophane Human - 100IUml in 10ml Vial – Injection (Suspension) and forecast of MSE.



The run where the Multi-Layer Perceptron predict for Dextrose - 40% in 20ml - Intravenous Infusion and forecast of MSE.



# APPENDIX C

```
#Auto ARIMA Forecasting in R demand prediction

library(readxl)

library(forecast)

#data <- read_xlsx("C:/Users/Natking/Desktop/DATA FOR PAPER/H.OFFICE
CONSUMPTION/TS4  IMP 1/Atropinex.xlsx")

data <- read_xlsx("C:/Users/Natking/Desktop/DATA FOR PAPER/H.OFFICE
CONSUMPTION/TS4  IMP 1/Dextrose.xlsx")

#data <- read_xlsx("C:/Users/Natking/Desktop/DATA FOR PAPER/H.OFFICE
CONSUMPTION/TS4  IMP 1/Insuline.xlsx")

#data <- read_xlsx("C:/Users/Natking/Desktop/DATA FOR PAPER/H.OFFICE
CONSUMPTION/TS4  IMP 1/TAT.xlsx")

tsdata <- ts(data$'Consumption', frequency = 12, star = c(2016,1))

#Plot the time-series formatted data

plot(tsdata)

#use teh auto.arima() function to get the optimal auto arima model

autoarima1 <- auto.arima(tsdata)

#get the forecasted data for a period of 1 year

forecast1 <- forecast(autoarima1,h=12)

forecast1

#plot the forecasted data from the auto arima model

plot(forecast1)

#plot the resuduals over time to see congruence or variance

plot(forecast1$residuals)

#plot the resuduals (sample vs. theoretical)

qqnorm(forecast1$residuals)

acf(forecast1$residuals)

pacf(forecast1$residuals)
```

# APPENDIX D

```
# LSTM for demand prediction problem with regression
import numpy
import matplotlib.pyplot as plt
from pandas import read_csv
import math
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import mean_squared_error
# convert an array of values into a dataset matrix
def create_dataset(dataset, look_back=1):
   dataX, dataY = [], []
   for i in range(len(dataset)-look_back-1):
      a = dataset[i:(i+look_back), 0]
      dataX.append(a)
      dataY.append(dataset[i + look_back, 0])
   return numpy.array(dataX), numpy.array(dataY)
# fix random seed for reproducibility
numpy.random.seed(7)
# load the dataset
dataframe = read_csv('C:/Users/Natking/Desktop/DATA FOR PAPER/H.OFFICE
CONSUMPTION/TS4 Missing Value Replaced 4/atropinex.csv', usecols=[1],
engine='python',
usecols=[1], engine='python',
skipfooter=1)
dataset = dataframe.values
dataset = dataset.astype('float32')
# normalize the dataset
```

```
scaler = MinMaxScaler(feature_range=(0, 1))
dataset = scaler.fit_transform(dataset)
# split into train and test sets
train_size = int(len(dataset) * 0.70)
test_size = len(dataset) - train_size
train, test = dataset[0:train_size,:], dataset[train_size:len(dataset),:]
# reshape into X=t and Y=t+1
look_back = 1
trainX, trainY = create_dataset(train, look_back)
testX, testY = create_dataset(test, look_back)
# reshape input to be [samples, time steps, features]
trainX = numpy.reshape(trainX, (trainX.shape[0], 1, trainX.shape[1]))
testX = numpy.reshape(testX, (testX.shape[0], 1, testX.shape[1]))
# create and fit the LSTM network
model = Sequential()
model.add(LSTM(4, input_shape=(1, look_back)))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
model.fit(trainX, trainY, epochs=100, batch_size=1, verbose=1)
model.compile(loss='binary_crossentropy', optimizer='adam',
metrics=['accuracy'])
#print(model.summary())
model.fit(trainX, trainY, validation_data=(testX, testY), epochs=100,
batch_size=64)
# make predictions
trainPredict = model.predict(trainX)
testPredict = model.predict(testX)
# invert predictions
trainPredict = scaler.inverse_transform(trainPredict)
trainY = scaler.inverse_transform([trainY])
testPredict = scaler.inverse_transform(testPredict)
```

```
testY = scaler.inverse_transform([testY])
# calculate Mean Square Error
trainScore = math.sqrt(mean_squared_error(trainY[0], trainPredict[:,0]))
print('Train Score: %.2f MSE' % (trainScore))
testScore = math.sqrt(mean_squared_error(testY[0], testPredict[:,0]))
print('Test Score: %.2f MSE' % (testScore))
# shift train predictions for plotting
trainPredictPlot = numpy.empty_like(dataset)
trainPredictPlot[:, :] = numpy.nan
trainPredictPlot[look_back:len(trainPredict)+look_back, :] = trainPredict
# PRINT PREDICTION
print(testPredictPlot)
print(testY)
```