



Strathmore
UNIVERSITY

SU+ @ Strathmore
University Library

Electronic Theses and Dissertations

2022

A Predictive analytics model for pharmaceutical inventory management.

Musimbi, Patience Musanga
Strathmore School of Computing and Engineering
Strathmore University

Recommended Citation

Musimbi, P. M. (2022). *A Predictive analytics model for pharmaceutical inventory management* [Strathmore University]. <http://hdl.handle.net/11071/13188>

Follow this and additional works at: <http://hdl.handle.net/11071/13188>

This work is available for free and open access by Strathmore University Library.
It has been accepted for digital distribution by an authorized administrator of SU+ @Strathmore University.
For more information, please contact library@strathmore.edu

A Predictive Analytics Model for Pharmaceutical Inventory Management



Master of Science in Information Technology

2022

A Predictive Analytics Model for Pharmaceutical Inventory Management

By

Musimbi Patience Musanga

136507

**Submitted in Partial Fulfillment of the Requirements for the Degree of Master of
Science in Information Technology at Strathmore University.**

School of Computing and Engineering Sciences

Strathmore University

Nairobi, Kenya.

October 2022

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Declaration and Approval

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

© No part of this thesis may be reproduced without the permission of the author and Strathmore University.

Student's Name: Musimbi Patience Musanga

Sign: _____ Date: _____

Approval

The thesis of Musimbi Patience Musanga was reviewed and approved for examination by the following:

Dr. Allan Omondi,

School of Computing & Engineering Sciences,

Strathmore University

Dr. Julius Butime.

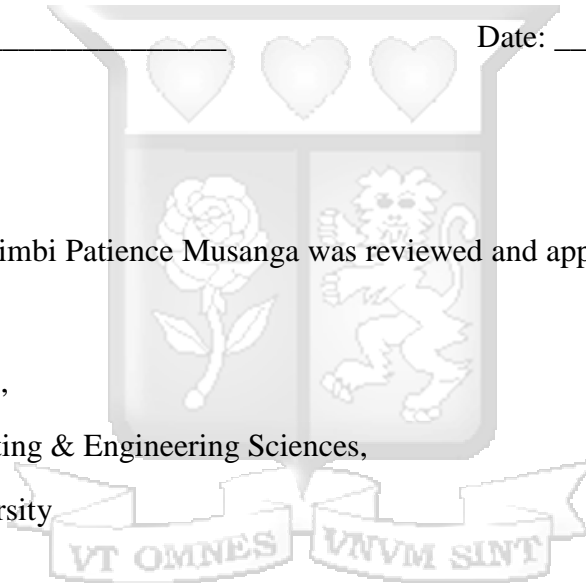
Dean, School of Computing & Engineering Sciences,

Strathmore University

Dr. Bernard Shibwabo,

Director of Graduate studies,

Strathmore University



Abstract

Inefficient inventory management is a factor that affects pharmacies in Kenya. The unpredictable nature of weather patterns during the traditional long and short rain seasons has resulted in seasons starting earlier or later than expected. Seasonal diseases such as flu may spike up when the temperatures decrease or when the rainy seasons begin, causing an increase in sales of drugs that cure and prevent the flu and vice versa. Due to this unpredictability, pharmacies may fail to stock up or down for different seasons due to unpreparedness and not knowing what to stock and when to stock. Ineffective drug management has a significant financial impact on pharmacies.

Inventory management ensures that needed drugs or medicines are always available, in sufficient quantities, of the right type and quality, and are used rationally. An effective drug management process ensures the availability of drugs in the right type and amount in accordance with needs, thereby avoiding drug shortages and excesses.

This research proposed a predictive analysis tool that would predict the required drugs or medicines prior to when they are needed, based on sales and seasonality. Another parameter for predictive analysis for this research was the period of the year when a certain disease could be common. This research discussed stocking and inventory management of pharmaceutical products and how predictive analytics with machine learning algorithms could be applied to improve the inventory management process in a pharmacy's context.

The purpose of the study was to examine the inefficient stocking of medicines in pharmacies and use predictive analysis to predict future stock. It reviewed various previous methods used for pharmaceutical inventory management and proposed the SARIMAX model with time series analysis for stock prediction. The result was a model that predicted the quantity of drugs to be stocked for the next six weeks. The six-week prediction model had a Root Mean Squared Error (RMSE) of 5.5.

Key words: ARIMA model, Machine learning, inventory management, SARIMAX, time series.

Table of Contents

Declaration and Approval	ii
Abstract	iii
List of Figures	x
List of Tables	xi
List of Models	xii
Abbreviations/Acronyms	xiii
Operational Definition of Terms.....	xiv
Chapter 1: Introduction	1
1.1 Background to the study	1
1.2 Problem Statement	2
1.3 Objectives	3
1.3.1 General Objective	3
1.5 Scope and Limitation	3
1.6 Justification	4
Chapter 2: Literature Review	5
2.1 Introduction	5
2.2 Theoretical Framework	5
2.2.1 History of Predictive analytics	6
2.2.2 ARIMA	8
2.2.3 SARIMAX.....	9
2.3 Empirical Framework	9
2.3.1 Pharmacy Inventory Management in Kenya	9
2.3.2 Climate in Kenya	10
2.4 Previously used Models	11

2.4.1 Auto-Regressive Integrated Moving Average (ARIMA) and Long Short-Term Memory Models (LSTM)	11
2.4.2 Support Vector Machine (SVM) and Artificial Neural Networks (ANN)	12
2.5 Current Methods of Inventory Management.....	13
2.5.1 Perpetual Inventory Systems	13
2.5.2 Automatic Dispensing Systems	13
2.5.3 RFID and Barcode Technology	14
2.5.4 Other Related Methods	15
2.6 Medical Inventory Prediction Model.....	15
2.7 Conceptual framework.....	16
Chapter 3: Research Methodology.....	18
3.1 Introduction.....	18
3.2 Research Design.....	18
3.3 System Analysis and Design.....	18
3.4 Target Population.....	19
3.5 System Development	20
3.5.1 Data Collection	20
3.5.2 Data Analysis	21
3.5.3 Data Pre-processing	21
3.5.4 Model Training	21
3.5.5 Model Validation	21
3.5.6 Deployment	22
3.6 Research Quality	22
3.7 Ethical Considerations	22
Chapter 4: System Analysis, Designs and Architecture	23

4.1 Introduction.....	23
4.2 Data analysis	23
4.3 Requirement analysis	24
4.3.1 Functional Requirements	25
4.3.2 Non-Functional Requirements	25
4.4 System Architecture.....	26
4.5 System Design	26
4.5.1 Context Diagram	27
4.5.2 Data Flow Diagram	28
4.5.3 Entity relationship diagram	30
4.6 Wireframes.....	30
4.6.1 Log in page	30
4.6.2 Admin Dashboard	32
4.6.3 Medication Page.....	32
Chapter 5: System Implementation and Testing.....	34
5.1 Introduction.....	34
5.2 Development requirements	34
5.2.2 Hardware requirements	34
5.2.3 Software requirements	35
5.3 Model Architecture	35
5.4 Model Development.....	36
5.4.1 Dataset overview	36
5.4.2 Pre-processing	36
5.4.3 Correlation	37
5.4.4 Time series Decomposition	38

5.4.5 Stationarity analysis	39
5.4.6 Autocorrelation	40
5.5 Training the model	40
5.5.1 Time series Forecasting	40
5.6 ARIMA model	41
5.7 SARIMAX model	41
5.8 Validating the model.....	43
5.8.1 Model Performance Results	43
5.8.2 Root Mean square	43
5.8.3 Forecasting Inventory with deployed model	44
Chapter 6: Discussion	45
6.1 Introduction.....	45
6.2 Challenges faced in stocking pharmaceutical inventory.....	45
6.3 Previously Used Methods for Inventory Management.....	45
6.4 To identify how pharmacies currently stock.....	45
6.5 Design of the Model using Predictive Analytics	46
6.6 Validation of the Model	46
6.7 Advantages of the Developed Model.....	46
6.8 Research Contributions.....	46
6.9 Challenges Encountered.....	47
Chapter 7: Conclusion and Recommendations	48
7.1 Conclusion	48
7.2 Recommendations.....	48
7.3 Suggestions for Future Research	49
References.....	50

Appendices..... 55

Appendix A: Gantt Chart..... 55

Appendix B: Sample drug and weather data collected..... 56

Appendix C: Ethical Approval..... 58

Appendix D: Data Collection Approval 59

Appendix E: Sample code..... 60

Appendix F: Similarity index..... 62



Acknowledgements

I would like to express my sincere gratitude to the Lord God Almighty for good health, strength, time and grace to undertake this study. I would also like to sincerely thank my immediate supervisor, Dr. Allan Omondi for his constant support, patience and dedication throughout this study. His guidance and commitment were key to achieving the research objectives and this is greatly appreciated. I would also like to thank the thesis coordinator, Dr. Omwenga who assisted with coming up with timely milestones that helped in achieving the research objectives. My father, Mr. James Musanga also offered great support and he is greatly appreciated. Lastly, I would like to thank my colleagues and coursemates for their insight and critique during this research.



List of Figures

Figure 2.1: Proposed model	16
Figure 2.2: Conceptual framework of the study	17
Figure 3.1: Iterative Methodology	19
Figure 3.2: Sample Drug sales data collected	219
Figure 3.3: Sample weather data collected	219
Figure 4.1: System architecture	26
Figure 4.2: Context diagram	28
Figure 4.3: Level 1 Data Flow Diagram	29
Figure 4.4: Entity Relationship Diagram	30
Figure 4.5: Log in page	31
Figure 4.6: Admin Dashboard	32
Figure 4.7: Predicted medication page	33
Figure 5.1: Merged dataset	37
Figure 5.2: Visualization of data	38
Figure 5.3: Time series decomposition view	39
Figure 5.4: Autocorrelation	40
Figure 5.5: SARIMAX forecast	42
Figure 5.6: Zefcoln forecast	43
Figure 5.7: Accuracy report	43

List of Tables

Table 2.1: A comparison of prediction methods.....	15
Table 5.1: Hardware requirements.....	150
Table 5.2: Software requirements.....	151



List of Models

1. ARIMA model
2. SARIMAX model



Abbreviations/Acronyms

AI - Artificial Intelligence

ANN - Artificial Neural Network

ARIMA- Auto-Regressive Integrated Moving Average

BN - Bayesian Network

CART - Classification and Regression Trees

DFD- Data Flow Diagram

DSS - Decision Support System

LSTM - Long Short-Term Memory

ML - Machine Learning

MSE- Mean Square Error

NDC- National Drug Code

PPB - Pharmacy and Poisons Board

RFID - Radio-Frequency Identification

RMSE- Root mean squared error

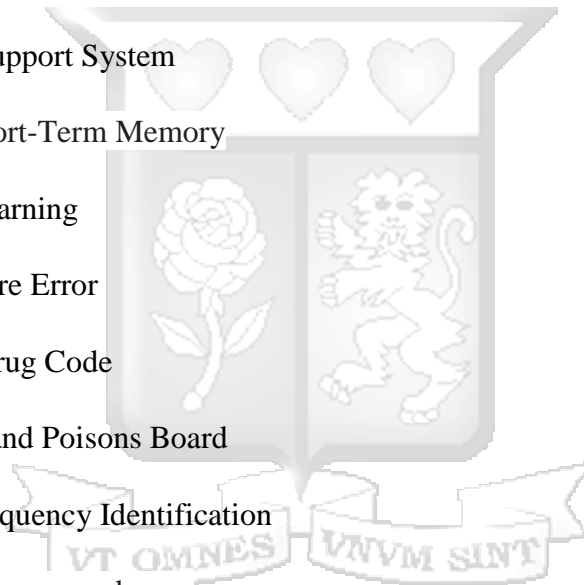
SARIMAX- Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors

SDLC - Software Development Life Cycle

SSAD- Structured System Analysis and Design

SVM - Support Vector Machine

WHO - World Health Organization



Operational Definition of Terms

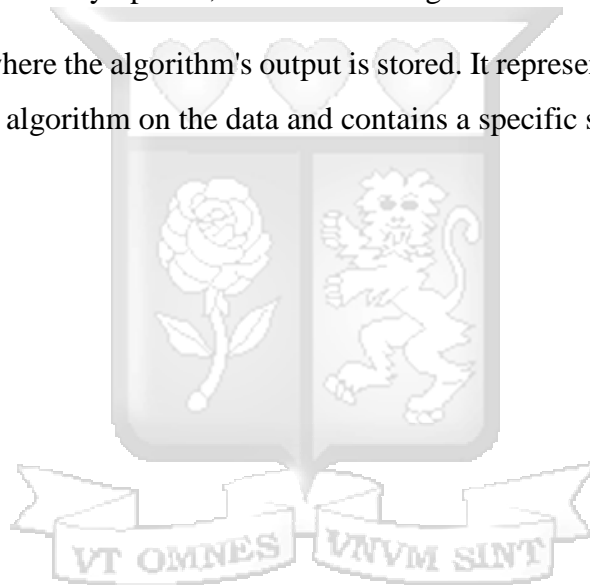
Algorithm - A procedure that is used to find patterns within data and learn from the data. They are implemented in code and are run on data.

Dispense – To prepare and distribute medication and other necessities to the sick, as well as to fill a medical prescription.

Forecast- The process of using historical data as input to make informed estimates for prediction.

Medicine/Drug - Medicines are chemicals or compounds that are used to treat, stop, or prevent disease, relieve symptoms, or aid in the diagnosis of illnesses.

Model – This is where the algorithm's output is stored. It represents what was learned from the training of the algorithm on the data and contains a specific set of algorithm features.



Chapter 1: Introduction

1.1 Background to the study

Drug management in pharmacies is an important component of hospital and pharmacy management. It is also the largest component that absorbs funding apart from medical services such as medical surgery and other types of treatment. According to WHO (WHO, 2015), good drug management constitutes rational selection, efficient quantification and forecasting, procurement, storage, and distribution of drugs. Drug management ensures that needed drugs or medicines are always available, in sufficient quantities, of the right type and quality, and are used rationally (International Finance Corporation, USAID, & IQVIA, 2020). Ineffective drug management has a significant financial impact on pharmacies. An effective drug management process ensures the availability of drugs in the right type and amount in accordance with needs, thereby avoiding drug shortages and excesses.

The pharmacy and poisons ACT CAP 244 laws of Kenya require the Pharmacy and Poisons Board (PPB) to regulate medical products and health technologies. All parties involved in the distribution of medical products and health technologies are responsible for ensuring that the quality of the products and the integrity of the distribution chain are maintained throughout the distribution process from the manufacturing to dispensing of the product to the end user (Pharmacy and Poisons Board, 2019). According to article 43 (1) (a) of Kenya's 2010 constitution, every person has the right to the best possible healthcare. These standards of health are only attainable if the quality of medical products and health technologies in the market are of the right quality and are dispensed correctly.

The aim of this paper was to shed light on the quantification and forecasting of medicine for stocking purposes and how predictive analytics with machine learning algorithms can be applied to improve this process. It is therefore worthwhile to define these concepts prior to the main discussion. Firstly, stocks refer to the quantity of finished products that are ready to be sold to an end user. Forecasting on the other hand is defined as the process of using historical data as input to make informed estimates for prediction. This process can be carried out in a clinic, hospital, health center or community pharmacy setting (Management Sciences of Health, 2012). While medicine administration is aimed at

improving patient care, lack of medication to administer to the patient can result in severe harm to a patient. All resources preceding it may go to waste if the patient does not end up receiving the right medication at the right time. This would increase the overall operating costs through wastage and shelving.

According to Mukherjee (2017), Artificial Intelligence (AI) and Machine Learning (ML) have been critical in the pharmaceutical industry and consumer healthcare business. For example, machine learning has been used to recognize diseases where it has been noticed that different types of medicines and drugs have been developed to treat cancer. Additionally, for a great analytical treatment, customized treatment is considered a more effective method in improving patient health outcomes because it is based on the health patterns. Finally, a plethora of studies speculate a widespread increase in the use of microdevices and biosensors, creating an opportunity for enhanced diagnosis, monitoring, and treatment of numerous diseases. The use of microdevices and biosensors promises to harness a huge amount of patient data, drug trends and seasonality of diseases. Unfortunately, knowledge and tools lack information and timely insights from collected big data. Thus, predictive analytics comes into play in analyzing the data to form medical insights resulting in stock efficiency in drug management.

1.2 Problem Statement

The provision of the right medication at the right time is a cornerstone of patient well-being. With the rise of prevalent diseases in different areas at different seasons, most pharmacies lack the ability to stock their shelves prior to the patients' needs due to the unknown demand for the drugs. Additionally, stocking without insight on what drugs are needed leads to understocking of the necessary drugs or overstocking of unpopular drugs (Okoneko, Khrustsky, Veber, Egorova, & Antropova, 2019). Consequently, understocking leads to lack of the required drugs while overstocking leads to excess inventory which may lead to expiry of the excess drugs. This lowers the customer(patient) satisfaction rate, unwanted inventory costs caused by overstocking and may lead to patients being harmed or worse, death. Therefore, using predictive analytics to gain insight on the seasonality

patterns and drug sales to predict what kinds of drugs to stock at a given time would be essential for both pharmacies and patients.

1.3 Objectives

1.3.1 General Objective

The purpose of this study was to apply predictive analytics to predict how to fine tune inventory management by monitoring drug sales and seasonal variation and advising on how to better stock a certain pharmaceutical product in order to control diseases.

1.3.2 Specific Objectives

- (i) To investigate the challenges faced in stocking pharmaceutical inventory.
- (ii) To review the previously used methods for stocking pharmaceutical inventory.
- (iii) To identify how pharmacies currently stock their inventory to best suit the needs of their customers.
- (iv) To design a model to predict future pharmaceutical inventory using predictive analytics.
- (v) To validate the designed model.

1.4 Research Questions

- (i) What challenges do pharmacies currently face while stocking their inventory?
- (ii) What methods for stocking pharmaceutical inventory have been used previously?
- (iii) How do pharmacies currently stock their inventory to best suit the needs of their customers?
- (iv) How can the proposed model predict future pharmaceutical inventory using predictive analytics be designed?
- (v) How can the designed model be validated?

1.5 Scope and Limitation

This study aimed at finding a solution using predictive analytics, that could impact on enhancing inventory management in pharmacies. Data collected for this research was collected from local pharmacies and online repositories and was limited to the number of

sales per day and weather data for each day. As discussed, the research would assist in inventory management, in the context of the pharmaceutical sector. Reasons for collecting data online included cutting back research costs and time. However, focusing on this scope had its setbacks. The research seemed to only apply to areas characterized by high number of traffic and unprecedented demand. Therefore, the application of the findings may be less applicable to pharmacies located in less busy regions.

1.6 Justification

This study examined the inefficiency in stocking of medicines in pharmacies, which was problematic. Using sales and weather data, it aimed at developing a model that would be used by pharmacists in medical institutions to categorize medicines mechanically and aid pharmacists in stocking and administering medicines appropriately. The system's target was to benefit pharmacies and pharmacists by predicting the quantity of drugs to stock so that pharmacies in future would neither be understocked nor overstocked, therefore maximizing profits. It also aimed to give a solution to aid in ensuring that patients receive the right medication at the right time, thus increasing customer satisfaction rate and of course less health complications or even deaths due to lack of drugs or administering expired drugs.



Chapter 2: Literature Review

2.1 Introduction

Inventory control has become an important component in supply chain management. One of the critical success factors in inventory management is accurate prediction. Many researchers have used different approaches to generate forecast of product demand for inventory control purpose. According to (Kerkkänen, Korpela, & Huiskonen, 2009), demand forecasting is commonly applied in companies that operate in consumer markets. Projections that are based on historical demand are typically very accurate when demand patterns are comparatively smooth and continuous. Success stories about demand forecasting typically report lower inventory levels and improved customer service.

In medicine, many problems have benefited from predictive analytics approaches. Large enough medical datasets have been available for a long time, but despite thousands of studies applying machine learning algorithms on medical data being done, very few have made a meaningful contribution to pharmaceutical care. This chapter provided a literature review of what had been previously used in this research's context, which is predictive analysis for inventory management to improve pharmaceutical care. Section 2.2 discussed the predictive analysis, computational learning as a fundamental principle of machine learning with more emphasis on regression, machine learning and predictive analytics. Section 2.3 is the empirical literature which discussed the history of inventory management in Kenya and climatology in Kenya. Section 2.4 discussed the previously used methods and algorithms for inventory management while section 2.5 demonstrated different traditional methods in which medical inventory management has been conducted previously. Finally, section 2.6 gave a summary of the proposed model, how it would be tested, its implementation and section 2.7 presented the conceptual framework.

2.2 Theoretical Framework

Drug scarcity is a complex issue that affects every aspect of the health-care system. According to (Baumer, et al., 2015), drug shortages have a wide-ranging impact, with more than half of health-care practitioners believing that shortages have influenced practice and resulted in subpar patient care. Replacement of drugs in the absence of the required drugs may have a negative impact thus resulting in medication errors (WHO, 2015). Excess drug

stock, on the other hand, causes issues for hospitals in the sense that money is wasted, and the excess drugs expire and become unsuitable for human consumption. Procurement of drugs that is not based on patient needs results in drug stockpiles accumulating (Nursalam, Saafi, & Munir, 2020). If pharmacy management does not consider larger storage space, the drugs become damaged and expire due to inactivity. Regular orders in small quantities can be placed to reduce storage costs. It should be noted, however, that out of stock occurs because purchase costs outside of planning can be high due to the high value of drugs (Management science of health, 2012). Even though much research has been conducted on the planning of stocks in pharmacies, many pharmacies and medical institutions still use traditional methods, which results in improper planning. Improper planning results in budget waste, stagnation, and stockouts. There are traditional inventory systems that have been used and other previous inventory management systems. In this section, they are discussed as part of the empirical literature.

2.2.1 History of Predictive analytics

This research was based on computational learning theory, which seeks to comprehend the fundamental principles of learning as a computational process (Sally, 2010). This field seeks to understand, at a precise mathematical level, what capabilities and information are fundamentally required to successfully learn various types of tasks, as well as the basic algorithmic principles involved in training computers to learn from data and improve performance through feedback. This theory aids in the development of better automated learning methods as well as understanding fundamental issues in the learning process itself. One exciting aspect of computational learning theory is the development of algorithms that quickly learn even in the presence of a large amount of distracting information. There are studies showing intelligent ways of predicting inventory requirements that have been proposed, bearing in mind that there are different concepts to put in consideration when ordering and managing inventory. Branches of computational theory i.e., Predictive analytics, machine learning and regression were discussed then selected for this research.

Predictive analytics is the analysis of current and historical facts to determine the likelihood of future events using data and statistical techniques from data mining, predictive modelling, regression, and machine learning (Nysce, 2007; IBM, 2021). It is divided into

three disciplines. First, the predictive models which evaluate the likelihood that a specific unit in a different sample has a similar performance. Second, the descriptive model that establishes the relationships in the data required for classification and third, the decision model that relates the data, the decision, and the result of the forecast. It is forward-looking; hence it uses past events to anticipate the future. Although predictive analytics has been around for decades and has been used for various applications, it has recently begun to gain popularity since most businesses are employing it to get insights about the future, mainly due to advancement of technologies and dependency of data (Korn, 2011). Many industries including banking and finance, energy industry and even the government and public sector are using predictive analytics to gain insights for future use (Ukhalkar, 2018). This research focuses more on regression and machine learning techniques which have been described below.

Regression analysis is a method that estimates relationships among variables. It focuses on developing mathematical equations as a model for representing interactions between various variables. It is intended for continuous data with a normal distribution and is mostly used to determine specific factors such as price (Predictive Analytics: What it is and why it matters, 2021). Regression examines how the value of the dependent variable changes when the values of the independent variables change in a modelled relation (Armstrong, 2012). Regression analysis is mainly used for prediction and forecasting. In some situations, it is also used to infer causal relationships between dependent and independent variables as mentioned earlier. There is a wide variety of models of regression models that can be applied when carrying out predictive analytics. They include linear regression models, logistic regression model, duration analysis and Classification and Regression Trees (CART) and discrete choice models.

Machine learning is a term that is used to refer to automated detection of meaningful patterns in data (Shai & Ben-David, 2014). It is a kind of AI that allows a system to learn from data and improve through experience without the need for explicit programming (Hurwitz & Kirsch, 2018). It employs several algorithms that learn from data in order to improve, describe, and predict outcomes. Machine learning has become popular in performing predictive analytics thanks to its techniques that have outstanding performance

in handling large datasets and noisy data (Linda, Joseph, & Ed, 2021). This involves training algorithms and neural networks to analyze data and outputting findings. There are two types of learning. Supervised learning and unsupervised learning. In this context, supervised learning creates predictive models using data that contains the results being predicted while unsupervised learning does not use previously known labels to train its models (Julianna Delua, 2021). It employs descriptive statistics to investigate the natural patterns and relationships that emerge from the data. Machine learning employs a variety of approaches, including Decision Tree Learning, Support Vector Machines (SVMs), Artificial Neural Networks (ANN), and Bayesian Networks, among others (Educba, 2021).

This research attempted to use predictive analytics to come up with a prediction of the required inventory. Time series approach was used for forecasting the future behavior of variables using time as an input parameter with the Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors (SARIMAX) model that is well suited for prediction of the value of an independent variable according to seasons was used.

2.2.2 ARIMA

The ARIMA model is characterized by three terms, i.e., p, d, q, where,

- (i) p is the order of the AR term- It refers to the number of lags of Y to be used as predictors
- (ii) q is the order of the MA term- the number of lagged forecast errors that should go into the ARIMA Model
- (iii) d is the number of differencing required to make the time series stationary- the minimum number of differencing needed to make the series stationary

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \epsilon_t + \Phi_1 \epsilon_{t-1} + \Phi_2 \epsilon_{t-2} + \dots + \Phi_q \epsilon_{t-q}$$

2.2.3 SARIMAX

ARIMA however, doesn't use seasonal differencing. To employ seasonality, we use SARIMAX which uses seasonal differencing with exogenous variables, in other words, it uses external data in this case, weather data like amount of rainfall and humidity to forecast. Therefore, if we include external data, the model will respond much quicker to its affect than if we rely on the influence of lagging terms. SARIMAX formula is given by:

$$\Theta(L)p\theta(Ls)P\Delta d\Delta Dsy_t = \Phi(L)q\phi(Ls)Q\Delta d\Delta Dset + \sum_{j=1}^n \beta_j x_{jt}$$

2.3 Empirical Framework

2.3.1 Pharmacy Inventory Management in Kenya

Pharmacies are key players in providing access to medicines and other pharmaceutical products in Kenya (Toroitich, Dunford, Armitage, & Tanna, 2022). Their influence is reflected in the growing interest to include them in the provision of essential pharmaceutical products and services. Though the scope of pharmacies varies, it would usually include registered and unregistered pharmacies which are governed by regulations similar to those of other health service providers. Some of these regulations include personnel qualification, structural design features of the premises, provisions for enough and good medicine, good medicine storage and good dispensing practices. However, studies have shown. This is due to challenges they face while stocking inventory. Some of the challenges faced in stocking pharmacies to be poor regulatory compliers in Kenya (Wafula, Abuya, Amin, & Goodman, 2014). According to (Deidre, Karrar, & Jayasree, 2018), one of the reasons being the lack of mechanism to maintain the availability of pharmaceutical stock pharmaceutical inventory include but are not limited to:

- i. The need to maintain the availability of products which are essential to human health or life itself without overstocking.
- ii. The need to maximize sales.
- iii. Strict regulations of the value chain, from research and development to marketing.
- iv. Perishable medical materials.
- v. Fragile production processes.
- vi. Sometimes relying on very small number of supply options.

With these challenges, optimizing inventory becomes more difficult for pharmacies and pharmaceutical companies. Calculating how much of a product to stock and how often therefore becomes a very important step in optimizing inventory. Because medications and biologics are expensive to make and cannot be held for long periods of time, it would seem logical that pharmacies would benefit from a lean, just-in-time (JIT) inventory strategy. However, COVID-19, has illustrated the problems that supply chain uncertainty poses to that model. At the same time, the perishable nature of pharmaceuticals, as well as their loss of potency as effective dates approach, jeopardize resiliency methods that rely on keeping safety stocks on hand. Finding the correct balance for optimizing inventory based on the product portfolio as well as historical demand and shipment patterns will be a constant task. While stocking, most pharmacies are affected by the bullwhip effect where small fluctuations in demand at the retail level can cause progressively larger fluctuations in demand due to the high shipment values involved, perishability of the product, and often the urgency in medicines where they're needed quickly. Some of the causes are chronic, like infrastructure congestion or labor shortages, viruses, and change in weather patterns.

2.3.2 *Climate in Kenya*

Kenya has a variety of different types of climates, such as a tropical rainforest climate in the southwest and a tropical monsoon climate in central Kenya and the southeast. Most places in Kenya have a rainy season and a dry season which depends on the location (World Bank Group, 2022). Kenya's temperatures vary, with the highlands experiencing considerably cooler temperatures than the coastal and lowland regions. Kenya's average annual precipitation is typically 680 mm. In general, the warmest period is from February to March, while the coolest is from July to August. There is the long rains period from March to May, and the short rains period from October to December. In Nairobi, Highs hover around 23/24 °C (74/75 °F) in the coolest months (June, July and August) and around 27/28 °C (80/82 °F) in the warmest months (January, February and March), while lows drop to around 12/13 °C (54/55 °F) from June to September and go up to 14/15 °C (57/59 °F) from January to April. In July and August, the sky is often cloudy, even though there is little rain, and sometimes at night it can be even cold, in fact, the temperature can drop

to around 5 °C (41 °F). However, Kenya's climate is changing. Rainfall patterns have changed, with the long rainy season becoming shorter and dryer and the short rainy season longer and wetter (Government of the Republic of Kenya, 2018). Different climates may relate to different types of illnesses. For example, during the cold season, Flu is most common. According to (Emukule, et al., 2016), there are multiple flu epidemics occurring each year and lasting a median duration of 2 months, with the first epidemic occurring between the months of February and March and the second one between July and November. Humidity is independently and negatively associated with flu. Combinations of low temperature and low humidity are significantly associated with increased flu. This research combines climatical patterns, drug sales and dates to predict the quantity of drugs to be stocked in pharmacies to prevent or cure these illnesses.

2.4 Previously used Models

Existing researches in the field of pharmacology identify several most effective and accurate methods: a linear regression (LR), random forest (RF) method, construction of a time series prediction using a neural network (NN), Auto-regressive integrated moving average method (ARIMA), long short-term memory model (LSTM), the use of support vector regression (SVR) and the LevenbergMarquardt algorithm (LMA) **Invalid source specified..** A few of them are discussed in this section.

2.4.1 Auto-Regressive Integrated Moving Average (ARIMA) and Long Short-Term Memory Models (LSTM)

The Auto-Regressive Integrated Moving Average (ARIMA) model is a model for time series prediction which can capture a suite of different standard temporal structures in time-series data (Adam Hayes, 2021). It uses correlation between current observations and past observations. For example, (Matsumoto & Ikeda, 2015) conducted an examination of demand forecasting by time series for auto parts manufacturing using time series analysis of actual shipment data from an independent remanufacturer. (Fattah, Ezzine, Aman, Moussami, & Lachhab, 2018) also used historical demand information to develop several ARIMA models using Box-Jenkins time series procedure, to forecast future demand in food manufacturing. Furthermore, ARIMA and LSTM techniques establish rolling forecast

models, which significantly improve accuracy and efficiency of demand and inventory forecasting. The forecast models, developed through historical data, are evaluated, and verified by the root mean squares and average absolute error percentages in the actual case application (Wang, Chien, & J.C.Trappey, 2021). ARIMA and LSTM models predict the top five products and validate the actual data and prediction results with Root mean squared error (RMSE) to evaluate the prediction model's performance. Consequently, LSTM has the smallest forecast error in the short-term forecast. However, its disadvantage is that the time-series data must be stable after differentiation. Another disadvantage is that only linear relationships can be captured in essence and not nonlinear relationships (Lu, Chunxue, & Neal, 2022). It is also only suitable for short term predictions.

2.4.2 Support Vector Machine (SVM) and Artificial Neural Networks (ANN)

Artificial Neural Networks have previously been used, but they require large training datasets (Gutierrez, Solis, & Mukhopadhyay, 2008). The concept of estimating time series structural components across multiple frequencies and optimally extrapolating and combining them, with empirical results promising for long-term forecasting, was proposed (Kourentzes, Petropoulos, & Trapero, 2014). Artificial neural networks (ANN) have shown better performance in classification and regression issues in the pharmaceutical sector during the past few decades, and they have attracted a lot of interest in time series forecasting techniques. However, ANN had certain drawbacks, including a long development time and a large amount of data that was needed. Due to the frequent updating of medications and the dearth of data on historical sales of individual preparations, it is required to develop an effective model for predicting pharmaceutical sales using one of the machine learning techniques (Keny, Nair, Nandi, & Khachane, 2021). SVMs and neural networks were later proposed as forecasting method combinations with improved forecasting performance (Petropoulos, Nikolopoulos, Spithourakis, & Assimakopoulos, 2013). In the automobile industry, SVM has previously been used for demand forecasting of automobile parts (Agarwal & Jayant, 2019). Both neural networks (NNs) and support vector machines (SVMs) are common machine learning approaches with applications in prediction based on times series data. Neural Networks have been used successfully for

pattern classification and recognition, weather forecasting, data mining and knowledge discovery, and in time series prediction tasks such as financial market prediction, stock prices and foreign exchange forecasting (Lucas & James, 2010). However, they both have disadvantages. For SVMs, minor fluctuations in training data causes decrease in predictive ability while for ANNs, the predictions become worse as the noise variation increases.

2.5 Current Methods of Inventory Management

2.5.1 Perpetual Inventory Systems

Perpetual inventory systems are systems that continuously record the quantity of a specific medication as prescriptions are filled through a point-of-sale system (Gupta, 2020). After each prescription is filled and dispensed to the patient, the medication used for the prescription is removed from the inventory to ensure that the quantity on hand in the computer is always current. Deliveries and returns are also recorded as they happen automatically.

Perpetual systems are designed to automatically update available quantities as prescriptions are filled, as well as to generate automated and manual reports that allow pharmacy staff to analyze and monitor inventory (Katie Ingersoll, 2017). They can frequently track turnover rates, predict future drug needs, alert pharmacy staff when potential errors are detected, and even automatically order more medication based on predefined reorder points. When medication levels are low, many pharmacies use periodic automatic replacement (PAR) levels in perpetual inventory systems to automatically order more medication. When a medication stock level in the perpetual inventory system is reduced to the pre-set minimum level, the computer system automatically orders enough medication to reach the maximum level, resulting in a simplified ordering system and reduced workload for pharmacy technicians.

2.5.2 Automatic Dispensing Systems

Using Automated pharmacy dispensing systems is also a method used for dispensing medicines. Automated pharmacy dispensing systems are management systems that allow

for storing and dispensing medicines near the point of use (Nicole, Clifford, Michele, & Kieran, 2014). They offer computer-controlled medication storage, dispensing, and tracking. The Pyxis and Omnicell machines, for example, are commonly used in hospitals to maintain stock of prescription medications to assist patients with their medication needs. To improve the efficiency of medication distribution, automated pharmacy dispensing systems have been recommended. They enable a more streamlined medication dispensing system while also increasing the pharmacy's ability to track system users and the items they add or remove, as well as provide reports on which drugs need to be refilled in the cabinets. They provide secure medication storage on patient care units as well as electronic tracking of controlled drugs (Ingersoll, 2015). However, their ability to reduce medication errors is dependent on a variety of factors, including how users design and implement the systems. To enhance and maintain pharmaceutical care evolution, automated dispensing should be improved (Berdot, et al., 2019). Other emerging technologies that will help hospital pharmacies' efficacy and possibly decrease the likelihood of adverse drug effects are recommended (Tsao, Lo, Babich, Bansback, & Shah, 2014).

2.5.3 RFID and Barcode Technology

Radio Frequency Identification (RFID) is a technology that allows objects to be tracked by connecting them to the internet. A bar code, on the other hand, is a method of representing data in a visual, machine-readable format. Many medications have barcodes on their packaging to facilitate product identification in a computer system. The barcode includes the product's National Drug Code (NDC) number, which tells the computer the product's name and package size. While barcode applications require line-of-sight identification, RFID tags are robust and do not require it (Peak Technologies, 2019). This technology contributes to the elimination of the need for human intervention. The technology uses programmable tags that contain information such as destination, weight, and a time stamp. RFID allows for warehouse space optimization and efficient goods tracking, which reduces costs and improves customer service. RFID tags can also communicate in real time and provide accurate information (Laquanda, Kamal, & Peebles, 2017). The use of RFID technology in the management of hospital supplies has the potential to significantly reduce

hospital inventory levels, as inventory is always a cost to any business. The main advantage of RFID technology is the ability to track goods in real time throughout the supply chain. Real-time delivery time tracking enables Just-in-Time (JIT) manufacturing and retailing. JIT assists hospital purchasing committees in making strategic decisions (Joseph, Joshin, & Kumar, 2013).

2.5.4 Other Related Methods

Different methods have been proposed regarding machine learning. Some that have been earlier used to predict market stocks. In table 2.1 below, there is a comparison of the prediction methods.

Table 2.1: A comparison of prediction methods

Method	Advantage	Disadvantage	Parameters used
Support vector machine for prediction	Difficult to lose accuracy even when applied to a sample from outside the training sample	Minor fluctuations in data causes decrease in predictive ability	Consumer investment, net revenue, net income, unemployment rate
Artificial neural network	Lower prediction error	The more the increased noise variation, the more the worse the prediction.	Stock closing price
Hidden Markov Model	Used for optimization	Evaluation, decoding and learning.	Stock market trend
ARIMA	Efficient and robust	Only suitable for short term predictions.	Stock price

2.6 Medical Inventory Prediction Model

Each year at least millions of people get unwell due to lack of medication, as estimated by the National Academy of Science (Kamalanabanand & Premkumar, 2018). Concerning this

problem, a model that finds the ideal decision variables that affect the target variable while parsing the relevant features has been proposed. In this study, both ARIMA and SARIMAX were computed and compared according to accuracy. The proposed algorithm adopted SARIMAX model and addressed the problem by showing how an algorithm learned from data can optimize large-scale data and come up with a prediction. It demonstrated the prevalence of data-driven AI, which can be used autonomously in purely data-driven systems or in collaboration with domain knowledge in hybrid systems.

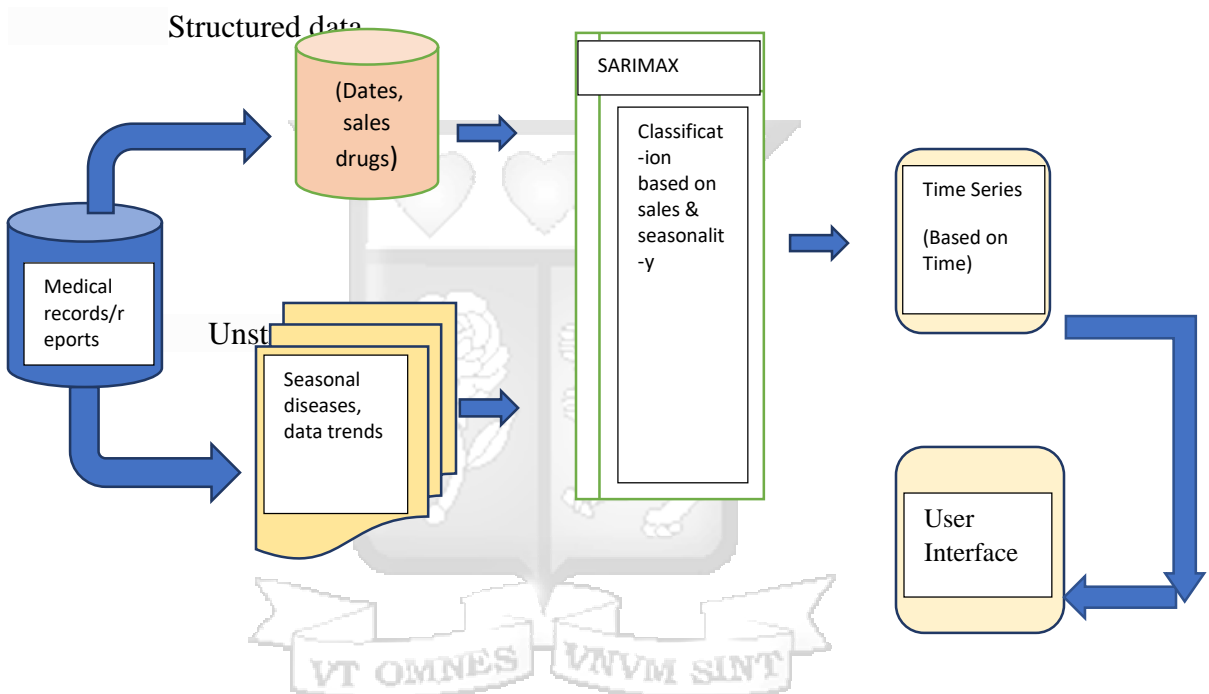


Figure 2.1: Proposed model

2.7 Conceptual framework

The study's aim was to apply predictive analytics to predict how to fine tune inventory management by monitoring drug sales and seasonality and advising on how to better stock a certain pharmaceutical product in order to control prevalent diseases. Most patients go to the pharmacy after attending the outpatient clinic while some are just walk-in customers. This model was expected to give the highest accuracy in comparison to other models. It would also help speed up patient waiting time for drug dispensing since the pharmacist would be expected to already have what they are requesting for. Changing the pharmacy

workflow would increase patient satisfaction and improve the overall quality of care to the patients.

In this model, the dependent variable is what the research is aiming to achieve. In this case, it is to enhance pharmaceutical inventory management. The independent variable on the other hand, is the predictive analytics which captures the machine learning and regression techniques, which is a correlation to the required output. These variables form the foundation upon which predictive analysis can begin. A series of tests would be done on the data set and the resulting performance of the algorithm will be computed. An analysis of these results will then be analyzed to see how the algorithm performs when applied to a data set. Figure 2.2: Conceptual framework of the study shows the conceptual framework of the model.

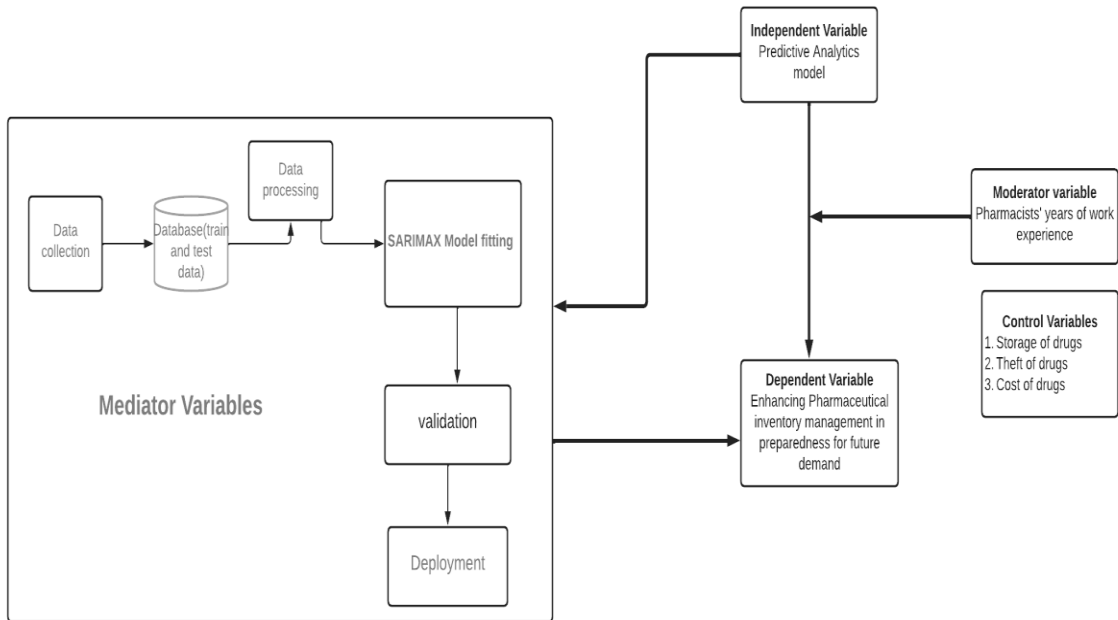


Figure 2.2: Conceptual framework of the study

Chapter 3: Research Methodology

3.1 Introduction

The research methodology that was used is outlined. The research design chosen and the population selected were also presented. This chapter covers the research design in section 3.2 and in section 3.3, it covers the system analysis and design. Section 3.4 describes the target population while section 3.5 gives a brief description of the model development process. Sections 3.6 and 3.7 present a discussion of the research quality and ethical approvals respectively.

3.2 Research Design

Research design is generally a framework on how various aspects of the research will be organized and conducted to combine relevance to the research objectives. This study combines both correlational research and applied design. This is because it gave the researcher the opportunity to describe the relationship between two measure variables (Cresewell, 2011), that is, if the weather related to the quantity of medicines purchased by patients at a certain period within the year. Correlational research was also selected because it is used to test the strength of association between variables and because other variables may play a role in the relationship. Conclusions can also be generalized to other populations or settings confidently with the use of this research design. This study is intended to help pharmacies receive stock quantity recommendations to best stock their pharmacies for future seasons.

3.3 System Analysis and Design

The final product of this study was a model that would be integrated in the pharmacy's current inventory system, that demonstrated key system functionality. For system analysis and design, Structured System Analysis and Design (SSAD) would be used since its focus was more on processes and procedures of the system. The design diagrams that were drawn included Data Flow Diagrams (DFDs), context diagrams and entity relationship diagrams. The approach selected to develop the application was the iterative methodology. The iterative methodology is used to design Software Development Life Cycle (SDLC) models that allow for creation of iterations that have the design, development, testing and review

phases. This means that each iteration would be reviewed in order to identify further requirements. This continued until the final product was achieved. Figure 3. shows the iterative methodology.

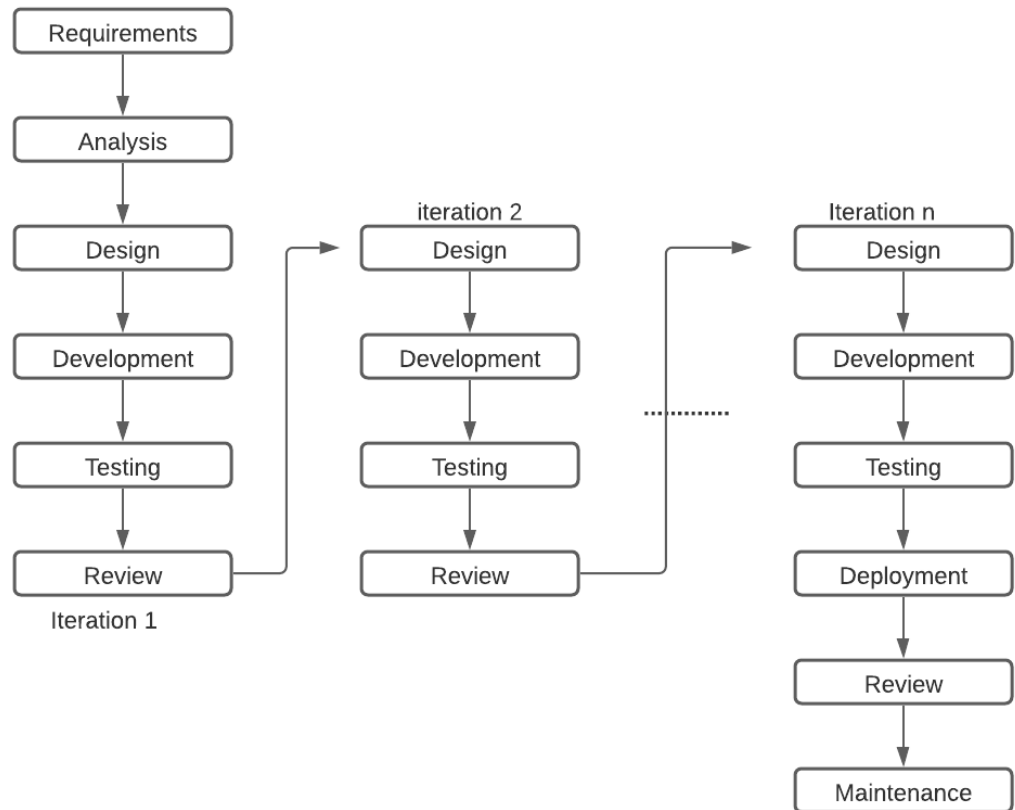


Figure 3.1: Iterative Methodology (Adapted From (Trivedi & Ashwani))

3.4 Target Population

This research targeted a study population constituting two pharmacies in Nairobi, Ruaka town. Data was also acquired from public repositories, i.e., Visualcrossing. Access to the data was requested while adhering to local requirements.

3.5 System Development

To develop the model, analysis of data collected was done. This process involved determining the factors influencing overstocking and understocking of pharmaceutical inventory in pharmacies. Data retrieved was cleaned and transformed to weekly time series consisting of cumulative sales among different pharmaceutical products. Data was analysed using python and graphically visualized to come up with inferences.

The system was developed using the below steps.

- i) Data collection
- ii) Data pre-processing
- iii) Model training and fitting
- iv) Model validating
- v) Deployment

3.5.1 Data Collection

This paper utilized data from local pharmacies that contained drug sales data and an online repository that contained weather data. Drug sales data was collected from two pharmacies in Ruaka town. Both have been in operation for the last 10 years. A brief discussion with the owners of the pharmacies who were both qualified and experienced proved that there was a defined pattern with the sales data at different times of the year. For instance, in July in all the years, there was the cold season with a high number of flu infections in different years. Data that was extracted from the drug dataset included the date of drug purchase, drug name and total quantity sold per drug. The dataset contained 994 records of drugs that were purchased for different illnesses for a period of 3 years. Data extracted from the weather dataset included the date, average temperature, average precipitation(rainfall) and humidity. The online repository mentioned above was Visual Crossing, <https://www.visualcrossing.com/> where a corpus of over 1558 records was retrieved from a database of historical weather data from the year 2011 to the year 2013 for Nairobi city. This data was used to conduct data analysis for the purpose of this study. The accuracy of forecasts was determined by considering how well the model performed on new data.

3.5.2 Data Analysis

To get an in-depth understanding of the effectiveness of the already available models, analysis of data was done. This process involved determining the factors influencing effectiveness of the previous models and establishing the challenges in their implementation. It also included determining the factors influencing overstocking and understocking of pharmaceutical inventory in pharmacies and hospitals. Data was cleaned and transformed to weekly time series consisting of cumulative sales among different drugs. Data was analyzed using python and graphically visualized to come up with inferences. Data used in this research was quantitative data since it could be quantified or measured. Therefore, it was analyzed and presented using graphs to understand the findings clearly. The research made use of inferential statistics, i.e., it analyzed relationships between variables. It accounted for sampling errors and included assumptions made regarding population distribution parameters. Correlation tests investigated the relationship between the variables and estimated the magnitude of the relationship.

3.5.3 Data Pre-processing

This involved systematically searching and arranging the data collected in a clearly understood way. Data used in this research was quantitative data since it could be quantified or measured. Therefore, it was analyzed and presented using graphs to understand the findings clearly. The research made use of inferential statistics, i.e., it applied autocorrelation to analyze relationships between variables and make comparisons. It accounted for sampling errors and included assumptions made regarding population distribution parameters. Correlation tests investigated the relationship between the variables and estimated the magnitude of the relationship.

3.5.4 Model Training

The model was trained and fitted with training data using the parameters for the regular ARIMA model (p,d,q) , as well as the seasonal ARIMA model (p,d,q,s) , i.e., the order and the seasonal order.

3.5.5 Model Validation

Validation was done using Root Mean Squared Error (RMSE) which is defined as the residual squared difference between the predicted values and the actual. It was achieved

using structured walk-through where predicted outcomes were compared with observed outcomes

The formula of RMSE is given by:

$$RMSE = \sqrt{[\sum(P_i - O_i)^2 / n]}$$

3.5.6 Deployment

The model was deployed using a web application developed using Python programming language while importing Flask as a library. The front end was developed using HTML 5 and Javascript.

3.6 Research Quality

The system was subjected to functional and non-functional tests, compatibility, and integration tests to determine whether everything worked within the stipulated requirements. Based on the feedback received, the model was then improved accordingly until the required product was realized. The research quality was measured in different dimensions that include integrity, inclusiveness, and relevance. It was considered as quality research once these dimensions were met.

3.7 Ethical Considerations

The data used was collected from pharmacies and open data repositories. This data is highly private and was treated with a high degree of confidentiality and solely used for the intended purposes. Additionally, all literature obtained from other sources such as journals, periodicals, books, etc., was referenced and cited aptly in this paper. Application for an ethical approval was made to the institution, Strathmore university, for institutional ethical approval.

Chapter 4: System Analysis, Designs and Architecture

4.1 Introduction

In this study, chapter four reports on the system analysis, design, and architecture. Considering the requirements collected in chapter three through the available datasets. This chapter clarifies the functionality of the developed system and the iteration between different components. Sections 4.2 gives more insight on what type of datasets were used and the functional and non-functional requirements of the system. Section 4.3 describes the system architecture. Sections 4.4 presents the system design with diagrams. Lastly, section 4.5 shows the system wireframes and how they will appear.

4.2 Data analysis

This involved systematically searching and arranging the data collected in a clearly understood way. Data used in this research was quantitative data since it could be quantified or measured. Therefore, it was analyzed and presented using line plots to understand the findings clearly. From the data collected, an analysis was done, and it was discovered that a high humidity and precipitation, led to more sales of flu drugs. Hence more people got flu during that period of the year, between the fifth month and the ninth month of the year. As depicted in the diagram, the total amount of drugs bought per week increased when there was high humidity and high precipitation. Pharmacists were then required to stock up during that period because of the high demand.

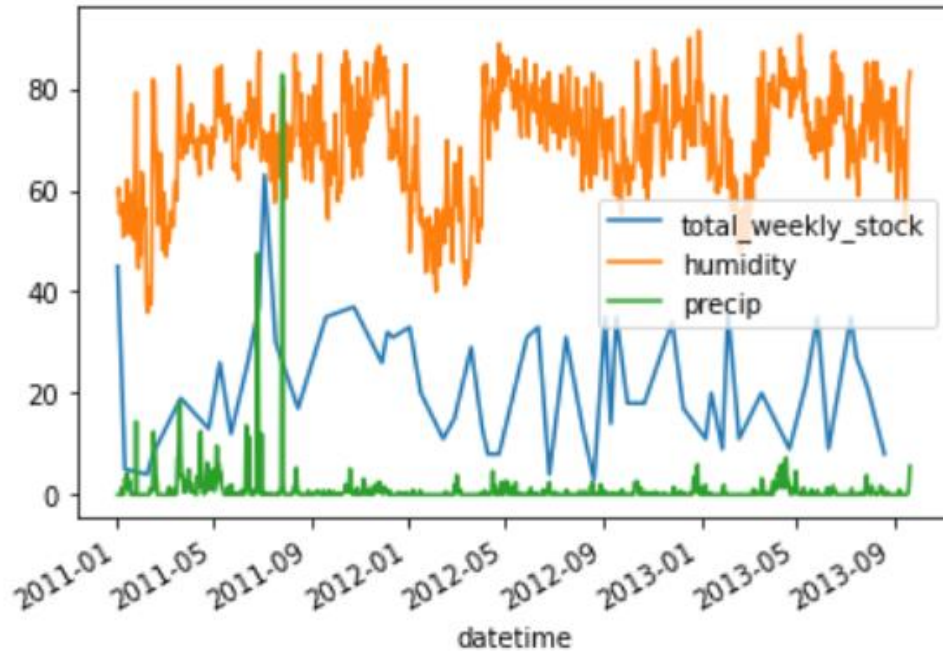


Figure 4.1: Data analysis

4.3 Requirement analysis

Requirement analysis involved the review of the functional, non-functional and operational requirements in order to ensure the model took into account all the stakeholders' needs as per the initial objectives of this study.

Initial requirements involved understanding the basics of the product requirements, especially the application's user interface in question. Considering the previous systems being used, the study intended to make the Graphical User Interface (GUI) more user-friendly to reach our goal. The system was expected to meet the functional requirements, meaning its actual performance from the end user's point of view would be as required. It was also required to meet the operational requirements of the organization, i.e., users' authentication and sign in, selecting the type of drug being forecasted for and loading of the csv data to be used for forecasting; hence it would speak to the needs of the organization, i.e., the need for pharmacies to be well stocked in preparedness for its customers. Technical requirements defined the technical needs, that is, the system would be installed in already existing equipment or a few new ones, thus reducing costs. Various

types of transition requirements, such as data conversion and migration, user access and security rights, user acceptance training, user preparation and transition, pilot testing, and infrastructure transition, were considered to meet the transitional requirements.

4.3.1 Functional Requirements

These are requirements that concern results or behaviours that are provided by a function of the system. They specify a function that a system or system component that must be made available to the users of the system. It must be independent of design and implementation aspects. The functional requirements are listed below.

- i. The system should allow for uploading of raw dataset as a csv file for training and testing.
- ii. The system should allow the admin to enter medicine factors on behalf of the pharmacist.
- iii. The system should provide drug inventory recommendations to be stocked based on previous sales and seasonal characteristics.
- iv. The system should allow for documentation of previous purchase records for future forecasting.

4.3.2 Non-Functional Requirements

Non-functional requirements define the desired qualities of the system to be developed and often influence the system architecture more than functional requirements do. They describe the non-behavioural aspects of a system, capturing the constraints under which the system must operate. The non-functional requirements are as follows:

- i. Availability- The system should be dependable, thus is expected to be functional around the clock.
- i. Transparency- The system shows how specific results were obtained to reduce issues with trust and transparency.
- ii. Security and privacy- The system should address privacy concerns when using the data acquired. It will also be safeguarded against deliberate and intrusive faults from both internal and external sources.
- iii. User friendly- The system should have a user-friendly interface, hence making it easier for learning and interacting with it.

- iv. Integrity- The system data should be maintained accurately and authentically, without corruption.
- v. Confidentiality- Data used should be private and confidential. The system will protect this sensitive data by allowing only authorised access to the data.
- vi. Efficiency- The system should be able to handle the capacity and throughput within the specified amount of time.

4.4 System Architecture

The system architecture for the predictive analytics model for pharmaceutical inventory management is shown below in Figure 4.. It explains the general interaction of various components to achieve system functionality. Raw data was pre-processed to create training and testing datasets. The predictive analytics model was then converted to a format that was to be embedded into the already existing system. The interface displays predictions as per users' requests.

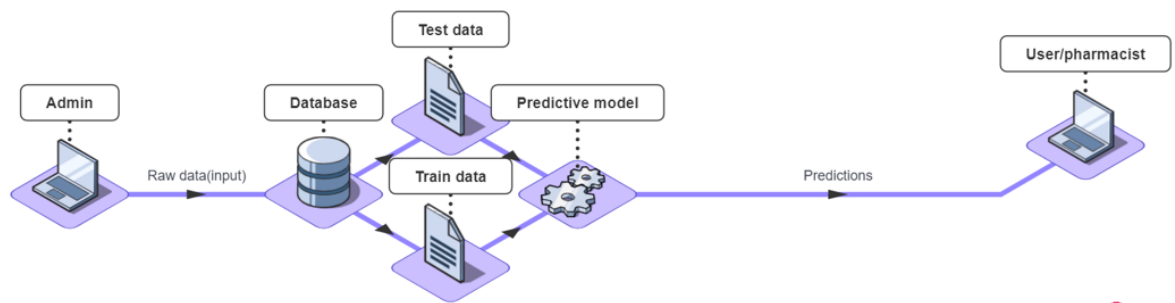


Figure 4.2: System architecture

4.5 System Design

The system design described how the system would be designed to meet inventory needs. The logical design pertained to an abstract representation of the data flow, the input, and the output. Physical design on the other hand, involved design of interfaces and processes to generate suitable specifications for the end product. To come up with the ultimate

system, the research was keen to follow the major tasks performed during the system design process, which included initializing the design definition to plan for and identify technologies that will implement the system's elements and their physical interfaces. Secondly, the study established design characteristics relating to the architectural characteristics. In addition to that, it assessed alternatives for obtaining system elements and manage the design.

To understand the model, various pictorial and graphical representations were used in the design stage. This section of this chapter therefore typically showcases the context diagram, data flow diagram and the entity relationship diagram.

4.5.1 Context Diagram

Context diagrams illustrate boundaries of the system, its environment and the entities that interact with it, that is the inputs and outputs from the system to its different entities. In the proposed model, the main entities that interact with the system are the user who is the pharmacist and administrator. The administrator maintains the required inventory prediction factors which are the previous sales data and the seasonal weather patterns. The user (pharmacist) enters the csv file that contains previous sales data and time of year. The model then calculates and makes a prediction on what quantities of medicines are required

to be stocked for the next six months. The administrator frequently updates the required factors for inventory prediction. Figure 4. illustrates the context diagram.

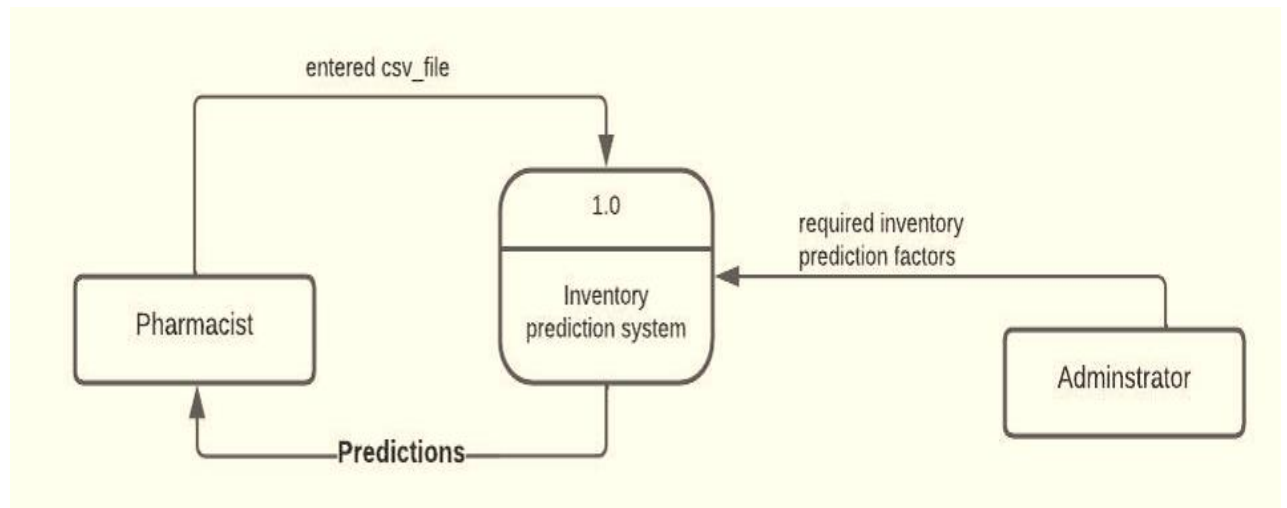


Figure 4.3: Context diagram

4.5.2 Data Flow Diagram

Processes and entities of a system demonstrating how data flows from the entities through the processes are illustrated by the data flow diagram. It captures the storage of data from the processes. The data flow diagram allows users to have a better understanding of the system.

The level 1 DFD gives a more detailed view by illustrating the various processes contained in the module, data stores and entities. Arrows depict the flow of data among various components of the DFD. Process 1.0 depicts an administrator adding medicine information together with factors affecting inventory into the system which is saved in the database. The user who is the pharmacist uploads a csv file which is also stored in the csv database for future use 2.0. The previous inventories are maintained as shown in process 3.0. In process 4.0, the administrator frequently updates the required factors that are required in determining the predictions of inventory. Feedback is sent to the user with information concerning the medicine quantities to be stocked from the pharmaceutical inventory prediction model. Figure 4. below presents the level 1 data flow diagram.

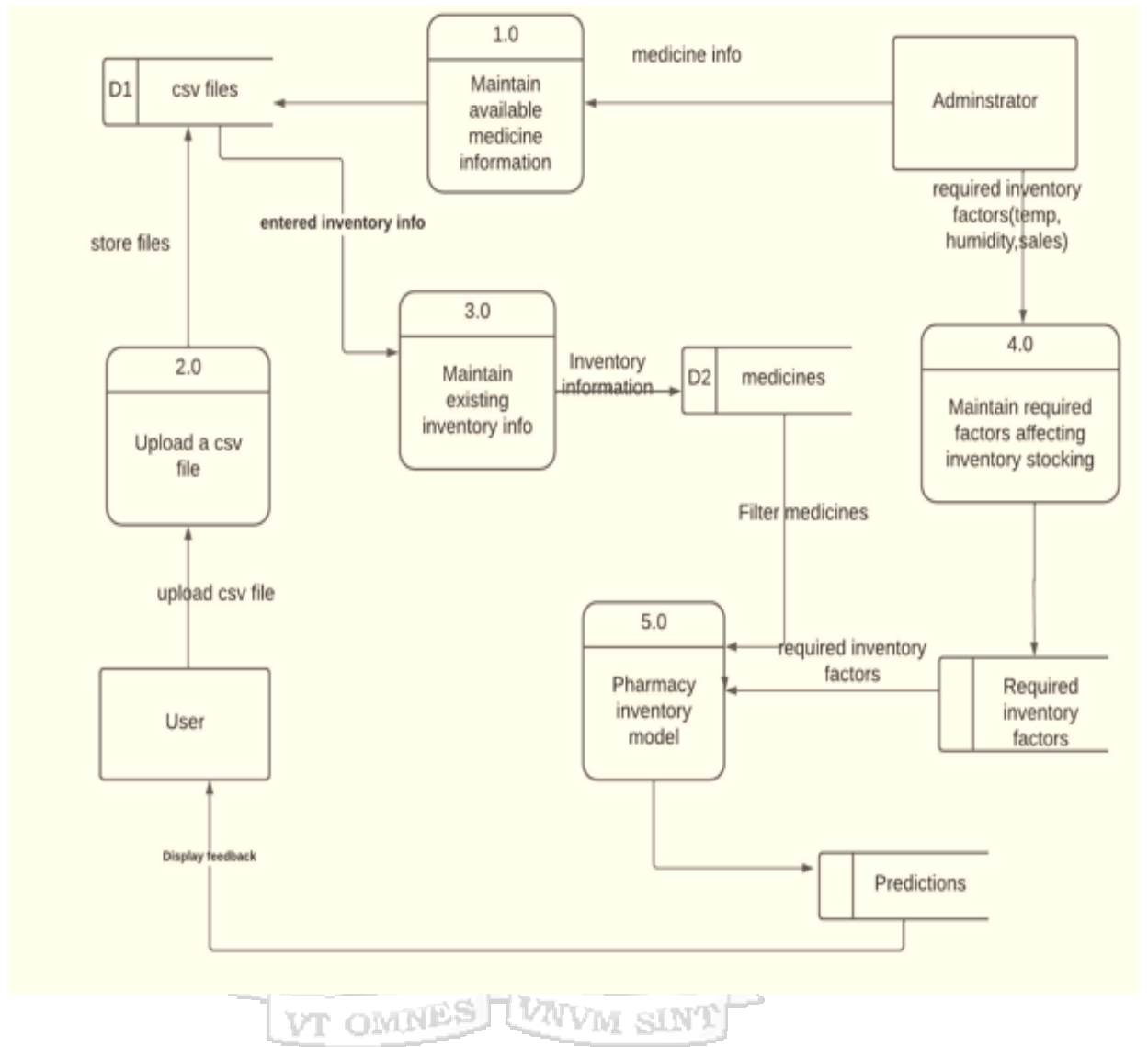


Figure 4.4: Level 1 Data Flow Diagram

4.5.3 Entity relationship diagram

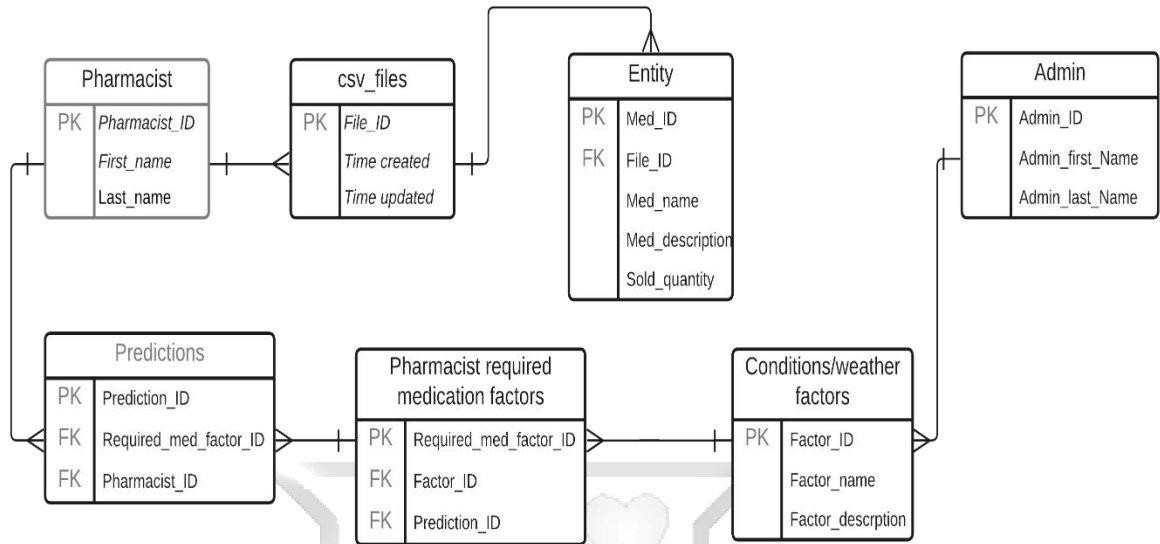


Figure 4.5: Entity relationship diagram

4.6 Wireframes

4.6.1 Log in page

This is the page where the user will sign into the system, indicating their name, employee number and entering their password. This will ensure that only authorised employees can log in to the system, meaning, only pharmacists with experience will be able to log in. If a user is not able to log in, they will be requested to sign up at the bottom. Figure 4. below shows the log in page.

Ipsium Pharmacy

Name

Employee No.

Password

I am a fully certified pharmacist with mandate to distribute medication.

Log in

Don't have an account yet? [Sign up!](#)

Figure 4.6: Log in page



4.6.2 Admin Dashboard

This is where sales data is represented graphically to show the performance of different categories at specific times and seasons, and people from what locations are buying what drugs. Figure 4. below show the admin dashboard.

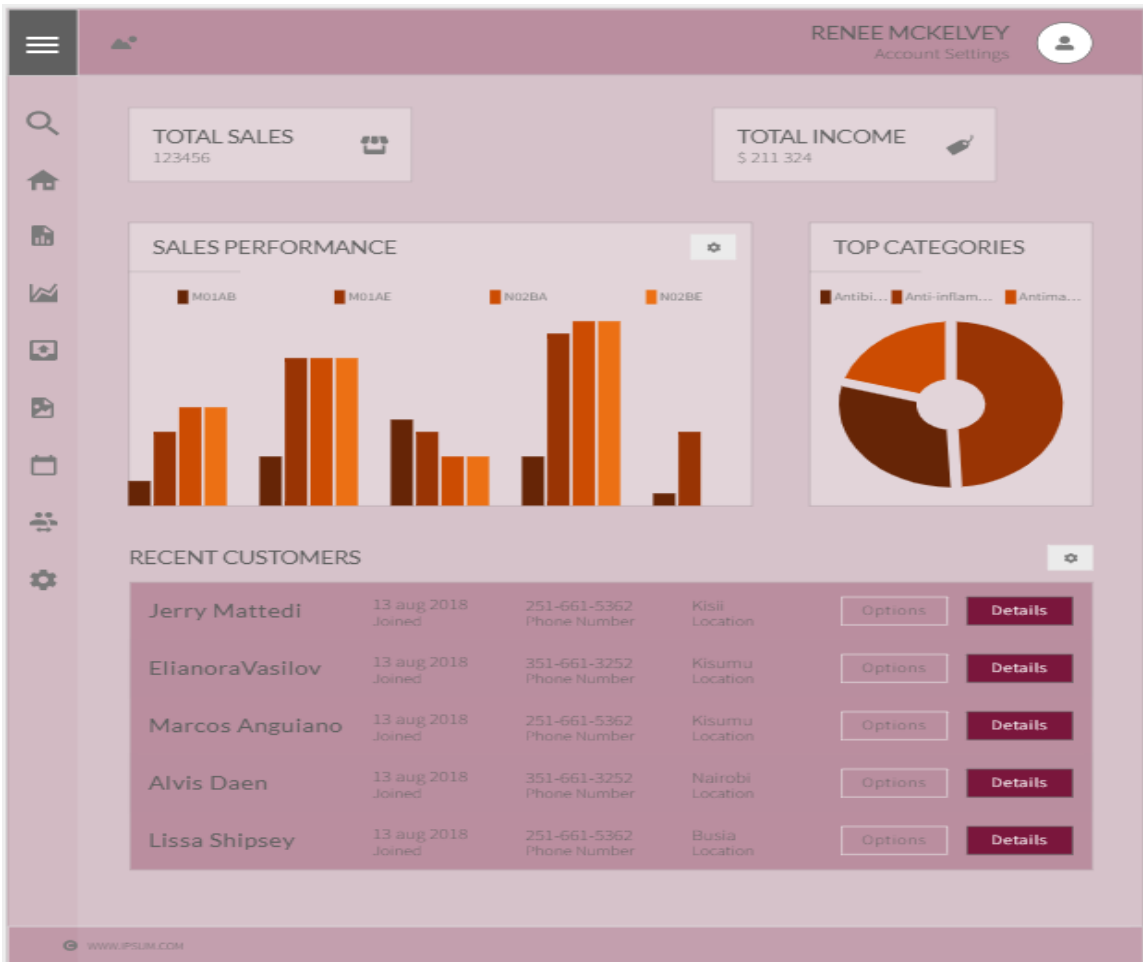


Figure 4.7: Admin Dashboard

4.6.3 Medication Page

Here, the illness or condition is entered, and the drug to be stocked for the illness is selected. Stock prediction is then displayed when the pharmacist clicks "Predict". Figure 4. displays the final stage of dispensing the medication.

Predicted Medication page

Name of illness/condition

Name of Drug

Upload your csv

No file chosen

Predicted Stock for the next six Weeks: 127.0 packets

Figure 4.8: Predicted medication page



Chapter 5: System Implementation and Testing

5.1 Introduction

This chapter focused on the implementation and testing of the model. The model development section, 5.2, discussed the hardware and software requirements for model development. Section 5.3 reviewed the systematic approach of how data was analysed and visualized. In section 5.4 to 5.6, the models were fitted and trained after which they were validated in section 5.7.

5.2 Development requirements

The model was developed using Jupyter Notebook which makes it easier to work on projects by allowing for running of code sections instead of running a whole python code. Python extensions were also installed. Python packages such as Plotly were used primarily in this study. A pre-processor was also used, which was adapted to medication data from other sources. To store data, MySQL was used because of its high performance and scalability.

5.2.2 Hardware requirements

Hardware Specifications	
Central Processing Unit (CPU)	Intel(R) Core (TM) i7-6820HQ CPU @2.70GHz (8 CPUs), ~2.7GHz
Memory	8GB RAM
Disk	256 Solid State Drive
Integrated Graphics chipset	Intel(R) HD Graphics 530

Table 5.1: Hardware requirements

5.2.3 Software requirements

Python 3.10.8 was used in programming since it is a general-purpose language and is primarily a language used for building machine learning models. Python is readable and has a good structure. It is also the most suitable for deep learning. Python Pandas was used for data processing since it is a tool for distributed computing and is faster and more convenient. It was used to enable faster data analysis and visualizations

Library	Version
Pandas	1.4.1
Numpy	1.22.3
Jupyter	1.00
Matplotlib	3.5.1
Flask	2.0.3
Plotly	5.9.0

Table 5.2: Software requirements

5.3 Model Architecture

The autoregressive moving-average (ARMA) model is used in ARIMA to evaluate and predict equally spaced univariate time series data, transfer function data, and intervention data. A value in a response time series is predicted by an ARIMA model as a linear combination of its own past values, past errors, and the present and past values of other time series.

In more detail, the terms AR give a description of the series based on its p past observations with auto-regressive coefficients, where p is the minimum number of prior observations required to predict the value of the series at the present. The MA component in a regression model displays a rolling average of past error terms rather than focusing on historical observations. Additionally, the integrated part is characterized by a parameter d , represents the order of differentiation.

5.4 Model Development

The model was developed using python language on Jupyter notebooks. Data was pre-processed and manipulated using Pandas library. Python was selected because it was best fit for machine learning program implementation and because it was easier to understand the python code. The sections below describe how the model was developed.

5.4.1 Dataset overview

Data retrieved was structured data. It had previously been used to conduct another research. However, it contained some information that was not relevant for the creation of the prediction model.

5.4.2 Pre-processing

The drug dataset contained data on the dates, drug name and quantity of drugs sold. During the pre-processing, some of these columns, including the condition and rating were dropped from the dataset. The weather dataset consisted of all climate data accumulated for 3 years. The drug dataset was then merged with the weather dataset to form one dataset that was then used for the prediction. The figure below shows a sample of the data contained in the final dataset. Seasonality had the biggest effect in terms of drug volumes; therefore, the sales column was decomposed estimate seasonal effects that were used to create and present seasonally adjusted values. Dates with missing values were dropped. A few types of drugs were then selected randomly to perform the forecast on. This included Centriline syrup, Coldcaps and Zefcoln.

	DrugName	Date	Quantity	datetime	temp	humidity	year	month	week
0	Flugone Capsules	2011-01-01	1	2011-01-01	21.3	59.3	2011	1	52
1	Goodmorning Syrup	2011-01-02	34	2011-01-02	20.6	60.4	2011	1	52
2	Rirathol	2011-01-03	29	2011-01-03	20.4	57.0	2011	1	1
3	Rirathol	2011-01-04	38	2011-01-04	20.8	55.2	2011	1	1
4	Centrizine 201	2011-01-05	24	2011-01-05	21.0	56.7	2011	1	1
5	Flugone Capsules	2011-01-06	23	2011-01-06	21.7	55.4	2011	1	1
6	Drtoux syrup	2011-01-07	26	2011-01-07	21.0	57.3	2011	1	1
7	Drtoux syrup	2011-01-08	1	2011-01-08	20.7	50.8	2011	1	1
8	Fungal dm	2011-01-09	21	2011-01-09	20.8	51.4	2011	1	1
9	Rirathol	2011-01-10	22	2011-01-10	22.1	54.1	2011	1	2

Figure 5.1: Merged dataset

5.4.3 Correlation

During visualization of the data, humidity data was plotted on a line graph showing humidity against days of the year. Another line graph was plotted to show how humidity and precipitation affects the patterns in which drugs are sold. As shown in the diagrams below, during the months between May and September, the total weekly stock sold seemed to go higher. High precipitation and humidity meant a higher purchase of drugs in pharmacies.

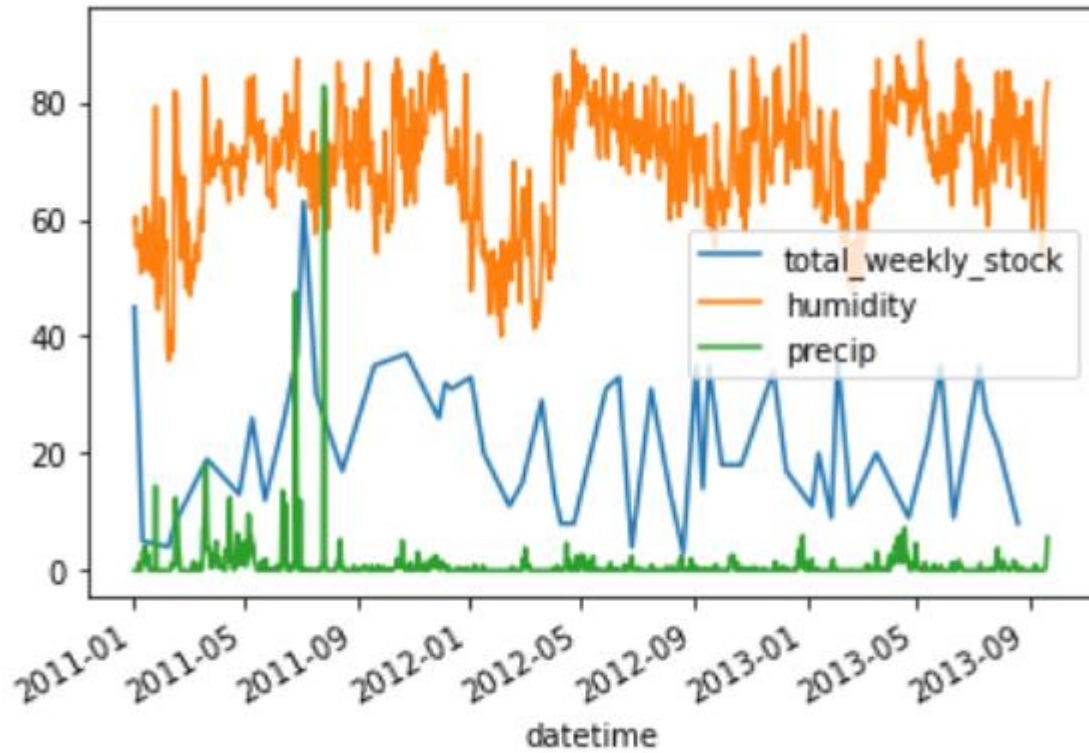


Figure 5.2: Correlation

5.4.4 Time series Decomposition

Trends and seasonality were explored in time series decomposition view. This was useful when determining uptake of residuals in data, based on the decomposed data. This was used to imply predictability since when decomposition is used, higher residuals always mean a lower predictability and vice versa. The figure below shows trends in the drug sales.

As the trend rises, so do the sales.

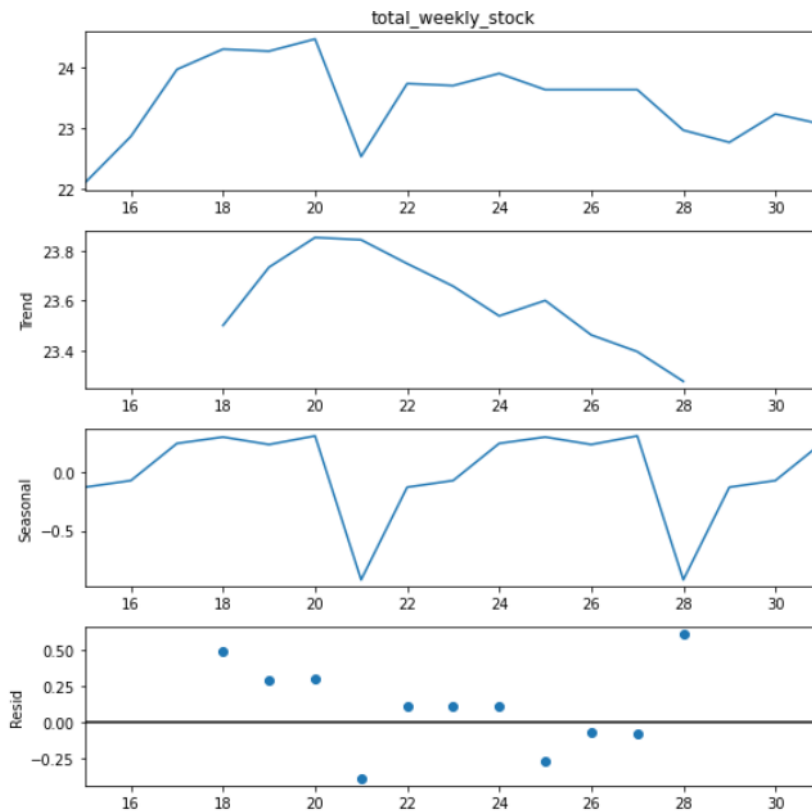


Figure 5.3: Time series decomposition view

5.4.5 Stationarity analysis

Data was calculated for stationarity using the Augmented Dickey-Fuller (ADF) test. The ADF test indicates the stationarity of a time series. Time series forecasting models like Vector Autoregressive model are dependent on time series stationarity hence need for the test. The p-value for Centrizine syrup and Zefcoln was not significant enough hence it was concluded that the time series was not stationary. It was however made stationary by differencing. Coldcaps time series were however stationary since the p-value was less than 0.05.

5.4.6 Autocorrelation

Autocorrelation was done for 'week', 'total weekly stock', 'average humidity' and 'average precipitation'. Autocorrelation analysis illustrates the potential for time series data prediction. It is used to summarize the strength of a relationship with an observation in a time series with observations at prior time steps. In the autocorrelation graphs, Count and humidity show a close autocorrelation

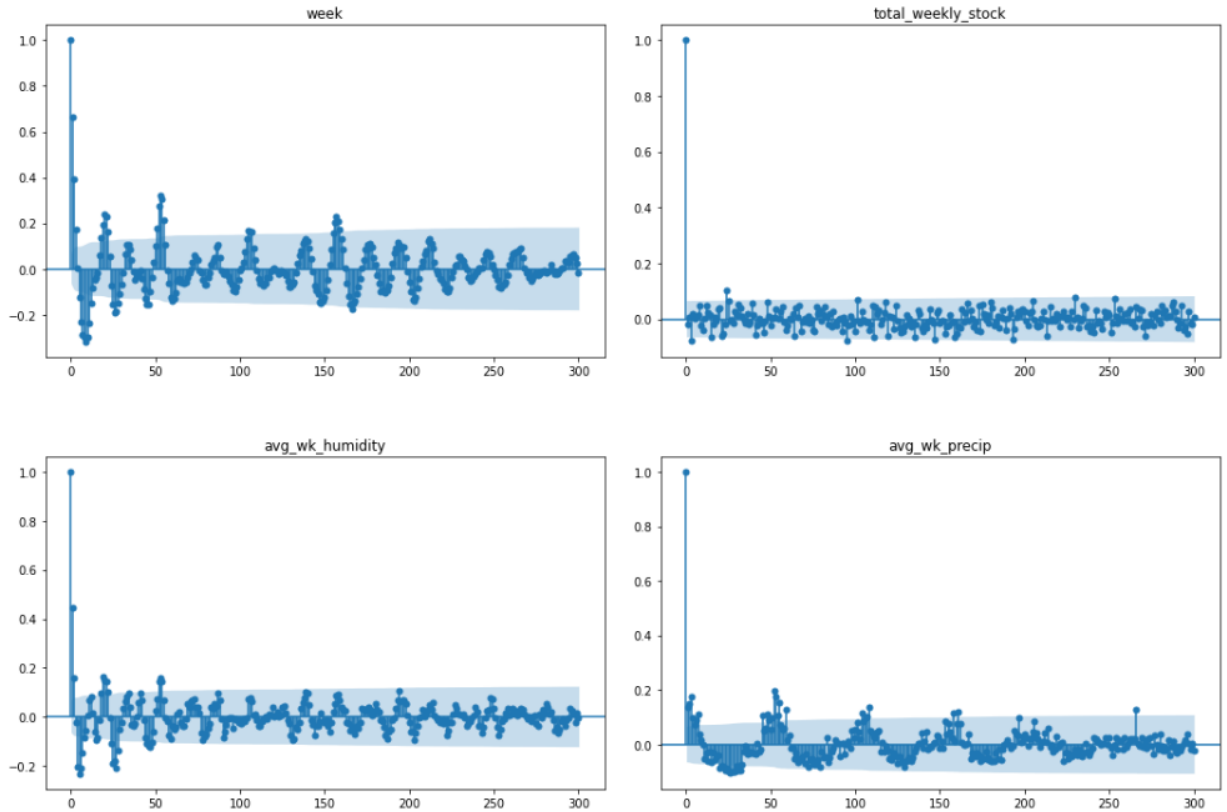


Figure 5.4: Autocorrelation

5.5 Training the model

The model was designed using the train data and tested using test data. This was done to identify stationary data. The corpus was first split into two with a ratio of 80:20. The training dataset which was 80% while the test data was 20%. Both SARIMAX and ARIMA models were fitted and compared using the Root Mean Squared Error (RMSE) to find out which one of the two had the best performance.

5.5.1 Time series Forecasting

ARIMA method was used to carry out short-term (rolling forecast) and long-term forecasting based on test data. Before each forecast was made, the process of optimizing hyper-parameters (p, d, q) of ARIMA model was carried out. Then, with optimal set of parameters, rolling forecast and long-term forecasting was carried out. Initial p and q parameters were first determined. SARIMAX was used since the dataset had seasonal cycles.

5.6 ARIMA model

The ARIMA model was fitted with p, d, q value of 1, 1, 1 respectively. It predicted total weekly stock. Figure 5.8 shows a graphical representation of how the forecast was made.

```
from statsmodels.tsa.arima.model import ARIMA
import statsmodels.api as smapi
model=ARIMA(Zefcoln_df[("total_weekly_stock")],order=(1,1,1))
history=model.fit()
```

Model 1: ARIMA model

5.7 SARIMAX model

The SARIMAX Model was fitted with p, d, q value of 1, 0, 1 respectively and a seasonal value of 1,1,1. Below is a figure that shows how the SARIMAX model was fitted.

```
#fit SARIMAX model
def SARIMAX_model(stock_vals,forecast_len,param_1):
    """
    Forecast using SARIMAX model
    """
    model= SARIMAX([("total_weekly_stock")], order=(1, 0,1), seasonal_order=(1, 1, 1,param_1)
                    ).fit(dis=-1)
    forecast_values_Zefcoln=model.get_forecast(steps=forecast_len)
    predicted=forecast_values_Zefcoln_df.summary_frame()["mean"].values
    return predicted
```

Model 2: SARIMAX model

SARIMAX predicted total weekly stock and gave an accuracy of 6.231343822062039. Figure 5.5 shows a graphical representation of how the forecast was made.

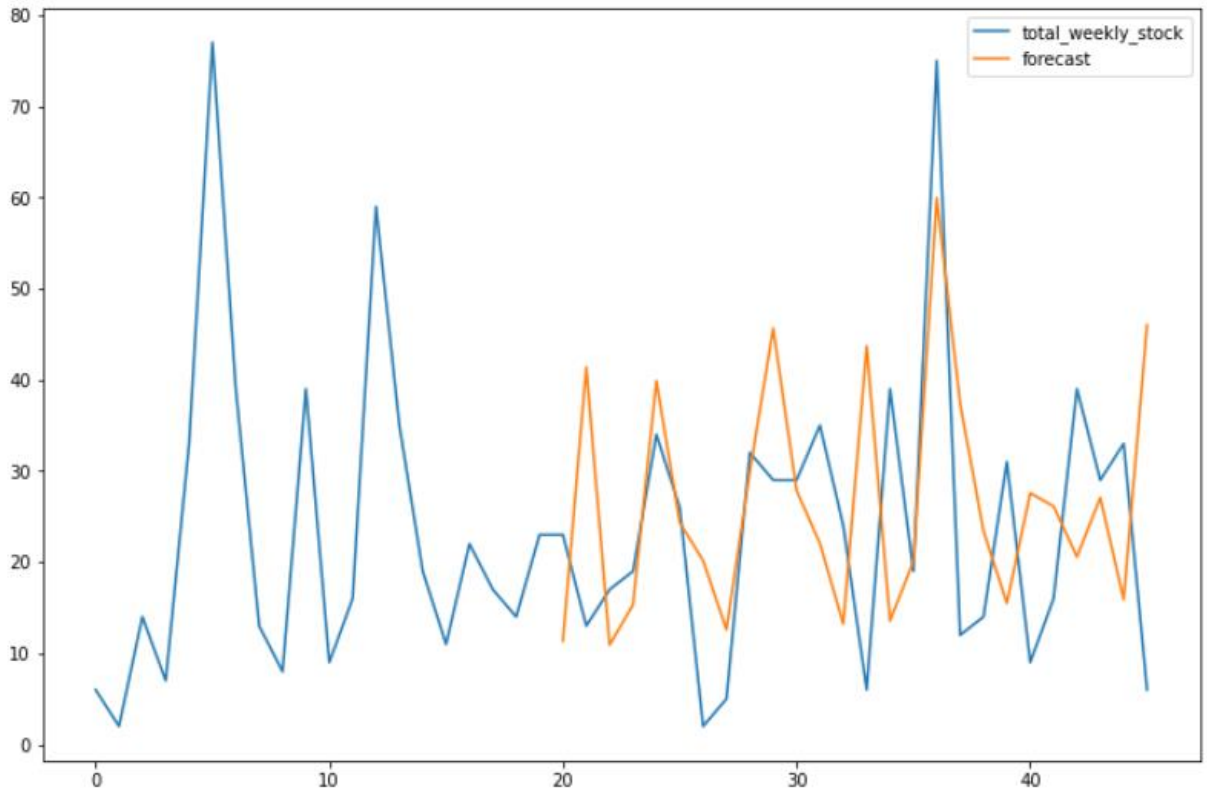
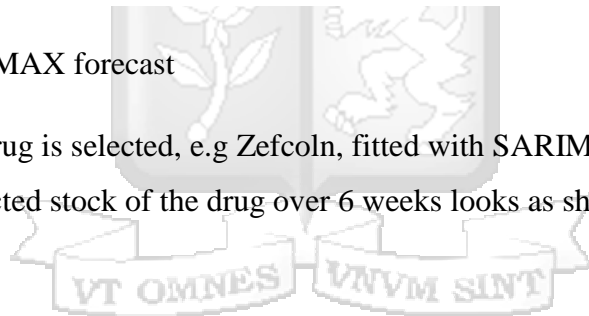


Figure 5.5: SARIMAX forecast

When a sample drug is selected, e.g Zefcoln, fitted with SARIMAX , the total weekly current and predicted stock of the drug over 6 weeks looks as shown in figure 5.6 below.



Zefcoln with actual and predicted

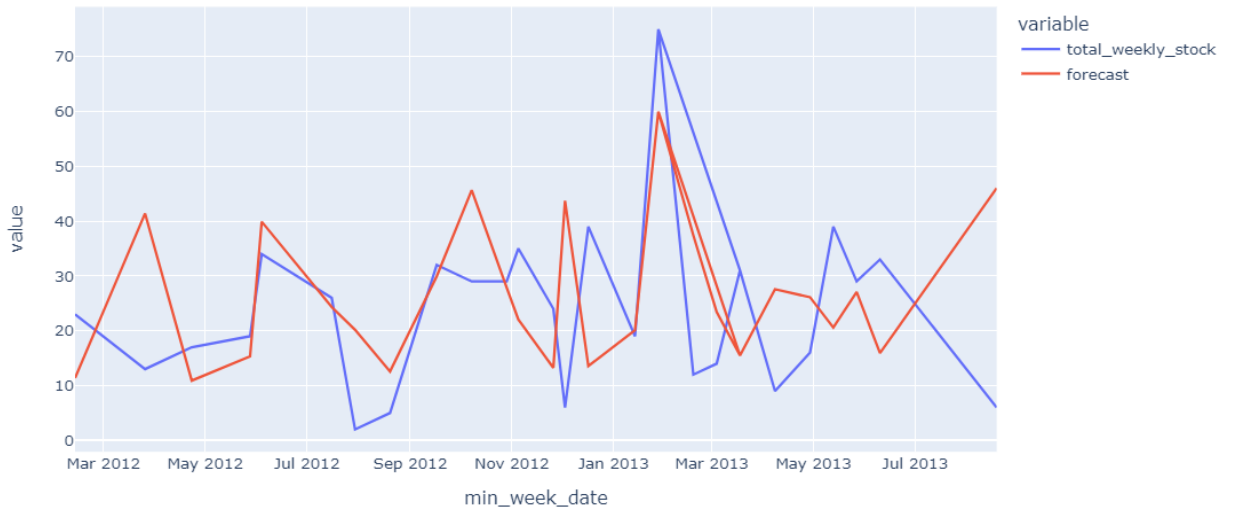


Figure 5.6: Zefcoln forecast

5.8 Validating the model

The testing data was used to validate the model. Model validation was conducted by assessing the error rates of the model based on validation data the model had not encountered during the training phase. The validation data consisted of approximately 20% of the collected dataset. The model was tasked to forecast the quantity of drugs for stocking in the next six weeks.

5.8.1 Model Performance Results

The experiment proceeded to determine the performance of each model by calculating the error rate of the models is determined by comparing the predicted drug quantity for stocking to the previous/expected drug quantity.

5.8.2 Root Mean square

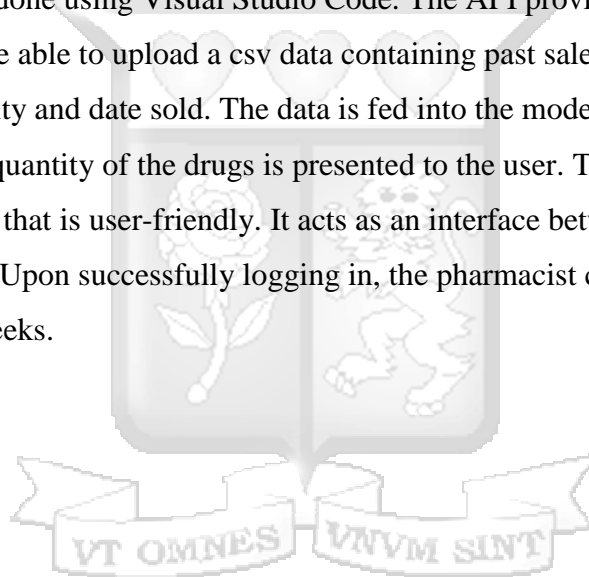
Accuracy for the model was tested using RMSE. RMSE is the residual squared difference between the predicted values and the actual. The RMSE value for the ARIMA model was 6.2 while that of SARIMAX was 5.5 which was a bit high owing to the fact the outliers were not removed from the data. The accuracy of the SARIMAX model was 88% as shown in the figure below, hence the model was good for prediction.

	precision	recall	f1-score	support
0	0.86	0.92	0.89	13
1	0.91	0.83	0.87	12
accuracy			0.88	25
macro avg	0.88	0.88	0.88	25
weighted avg	0.88	0.88	0.88	25

Figure 5.7: Accuracy report

5.8.3 Forecasting Inventory with deployed model

A backend Restful Application Programming Interface (API) was developed. The backend was developed using Python programming language while importing Flask as a library. This was done using Visual Studio Code. The API provides an interface to the model. One can be able to upload a csv data containing past sales data containing the drug name, quantity and date sold. The data is fed into the model and a six-week prediction of the quantity of the drugs is presented to the user. The frontend application is a web application that is user-friendly. It acts as an interface between the pharmacist and the API backend. Upon successfully logging in, the pharmacist can trigger the prediction for the coming weeks.



Chapter 6: Discussion

6.1 Introduction

This chapter discussed the results of the research in relation to the objectives. These discussions were a culmination of the tests and results obtained from the implementation and testing as presented in chapter 5. This research focused on the criteria used for stocking inventory for pharmacies. After reviewing different methods, the researcher chose the time series method which was best suited for seasonal data analysis and combined with SARIMAX forecasting.

6.2 Challenges faced in stocking pharmaceutical inventory

The first objective was to investigate the challenges faced in stocking pharmaceutical inventory. Based on the literature reviewed, the researcher found that optimizing inventory is difficult for pharmacies stating a myriad of challenges they face while stocking. The research also revealed that factors that affect stocking were seasons, trends of drugs and weather patterns among many others.

6.3 Previously Used Methods for Inventory Management

The second objective was to review the previously used methods i.e., ARIMA model (Adam Hayes, 2021), SVM (Gutierrez, Solis, & Mukhopadhyay, 2008) and ANN, and current methods such as the use of perpetual inventory systems (Gupta, 2020) and automatic dispensing systems (Nicole, Clifford, Michele, & Kieran, 2014). From the literature review, it was discovered that methods used to stock drugs only depended on if the drug is currently unavailable or if it needs restocking. Findings from the literature review also revealed that drugs could be reordered but not based on the trends of the drug, seasons, or prevalence of a disease, meaning they would still be understocked or overstocked.

6.4 To identify how pharmacies currently stock

The third objective was to determine what methods pharmacies currently use to plan for future inventory stocking. Based on the literature reviewed, the researcher found that most pharmacies stock inventories through guesswork and personal intuition. Some pharmacies

still use traditional methods, which results in improper planning, consequently resulting in budget waste, stagnation, and stockouts.

6.5 Design of the Model using Predictive Analytics

The fourth objective was to develop a prediction model for stocking pharmaceutical inventory. Research findings revealed that the developed model predicted drugs to be stocked six weeks ahead depending on the sales made and the seasonality of the drug.

6.6 Validation of the Model

The fifth objective was to validate the proposed model that predicts drugs to be stocked based on seasonality and sales. The validation was achieved using structured walk-through where predicted outcomes were compared with observed outcomes. Testing of the model was also done by uploading csv files containing sales and weather data of previous months and quantity of drugs to be stocked for the next six weeks was displayed. The drugs were determined by the season(weather) and the previous sales during that season in the previous years.

6.7 Advantages of the Developed Model

The most significant advantage of the developed model was that it used sales and weather data to predict, hence it was precise in giving insights. It also requested for input of the required drugs by the pharmacist so that this data could be stored for future purposes e.g., making refined predictions. The developed model required little human interaction as the pharmacist only interacted with it if they needed information on the quantity needed for a particular drug, while the administrator only interacted with it when entering the required factors to determine the prediction, i.e., weather patterns and previous drug sales. For the user interface, it gave a dashboard that shows all drugs for the condition entered and alerts for restocking or reorders of drugs depending on the users setting. The final product was a system that could be accessed through the web and was not dependent on operating systems on different devices. It could therefore be integrated with systems that run on the web.

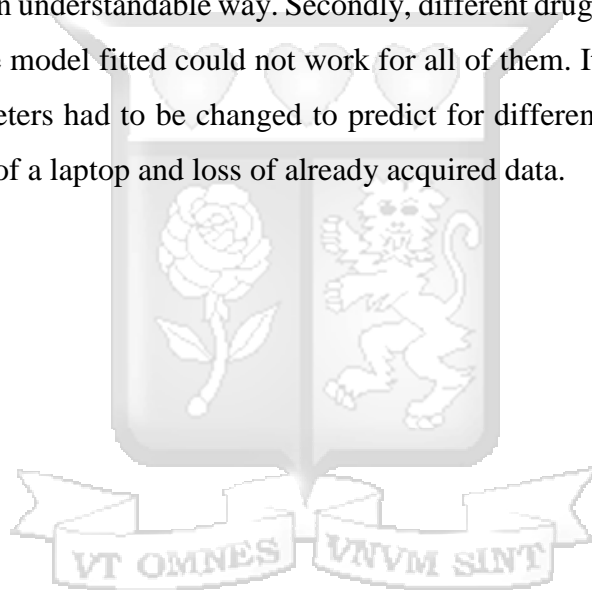
6.8 Research Contributions

The developed model provided a solution for estimating future stock of drugs in pharmacies. The model provided insights to pharmacists with regards to expected demand

for the next 6 weeks based on previous demands for that season for the previous years. The model could be used in pharmacies and hospitals to plan and budget for future stock to avoid wastage of resources, maximize on profits and help in ensuring optimal pharmaceutical care for patients.

6.9 Challenges Encountered

The most significant challenge was obtaining data. Most pharmacies and pharma companies do not disclose their sales information. However, I got two pharmacies to provide data. Some of the data obtained online lacked most of the factors needed or was too small to predict with. It was also time consuming going through the dataset as data was not structured in an understandable way. Secondly, different drugs had different seasonality patterns hence the model fitted could not work for all of them. It had to be fine-tuned i.e., the p, d, q parameters had to be changed to predict for different drugs. Other challenges included the loss of a laptop and loss of already acquired data.



Chapter 7: Conclusion and Recommendations

7.1 Conclusion

This research's main objective was to come up with a solution to solve the problem of stocking of pharmacies without knowledge of what and when to stock which is brought about by factors such as climatic conditions during different seasons. The research discussed models that had been used previously. The research also performed a literature review on previous research, shortcomings and implementations done in the past. This research focused on the sales and weather patterns to improve the decision-making process on stocking. The main deliverable was a model that predicts pharmaceutical inventory for stocking purposes. The final deliverable was a web application showing the prediction of a drug for the next six weeks. This deliverables' impact in the pharmaceutical industry includes the fact that it is much easier to navigate through the web application hence saving time while still making the right predictions in stocking. This involved the use of already existing datasets documenting daily drug purchases for the previous years and previous weather data as a basis for the forecasts. It also involved merging of the two datasets and splitting them into train and test data.

7.2 Recommendations

The model for prediction of inventory for stocking in pharmacies was a suitable solution for pharmacies. However, there were some recommendations since more could still be done in this area.

The following recommendation is made with regards to the research:

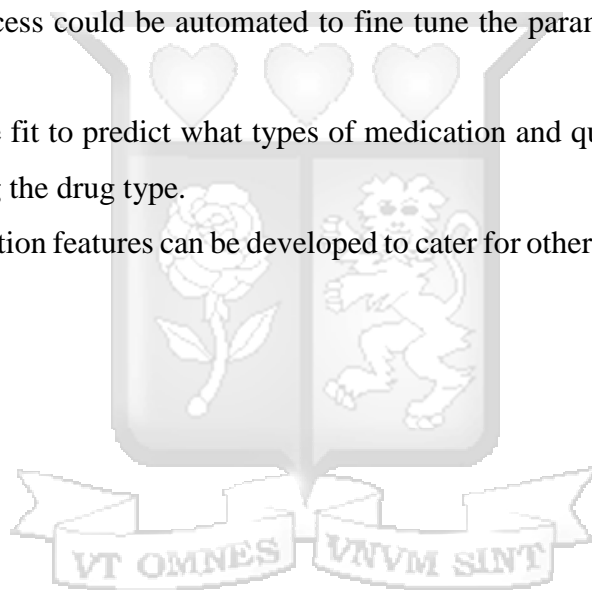
- i. Data was quite difficult to find. Recommendations were made to local pharmacies to keep records of their sales data and what methods they used before to stock inventory to allow for improvement of the model.
- ii. Pharmacies should be able to accept the model to be integrated into their current systems to allow for easy inventory management.

- iii. The model can be fit to predict for more or less than six weeks in preparedness for future demand, therefore, other retail businesses can also use the developed model for their inventory management.

7.3 Suggestions for Future Research

Future research suggestions include the following:

- i. The research couldn't include drug trends. Therefore, tweets could be crawled from twitter to get the current trends of a drug instead of only relying on the entered data. This will help strengthen the reliability of the model.
- ii. The model parameters couldn't be automated for this research. However, in future the model fitting process could be automated to fine tune the parameters used for each drug being forecasted.
- iii. The model can be fit to predict what types of medication and quantity to stock instead of manually entering the drug type.
- iv. More web application features can be developed to cater for other pharmacy functionalities.



References

- Government of the Republic of Kenya. (2018). *National Climate Change Action Plan 2018-2022*. Nairobi: Ministry of Environment and Forestry.
- Adam Hayes. (2021, 10 12). Autoregressive Integrated Moving Average (ARIMA). *Investopedia*.
- Agarwal, A., & Jayant, A. (2019). Support Vector Machine Model for Demand Forecasting in Automobile parts industry. *Research Journal of Applied Sciences, Engineering and Technology*, 33-49.
- Armstrong, J. S. (2012). Illusions in Regression analysis. *International Journal of Forecasting*, 689-694.
- Baumer, A. M., Clark, A. M., Witmer, D. R., Geize, S. B., Vermeulen, L. C., & Deffenbaugh, H. J. (2015). National survey of the impact of drugs shortages in acute care hospitals. *Pubmed*.
- Berdot, S., Blanc C, Chevalier , D., Bezie, Y., Le, L., & Sabatier, B. (2019). Impact of drug storage system: a quasi-experimental study with and without an automated-drug dispensing cabinet. *International Journal for Quality in Health Care*, 225-230.
- Cresewell, J. (2011). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Pearson.
- Deidre, C., Karrar, K., & Jayasree, K. I. (2018). Shortages, stockouts and scarcity: Issues facing the security of antibiotic supply and the role for pharmaceutical companies. *Access to medicine Foundation*.
- Doctors and Health practitioners in Nairobi, K. (2021). *Allianz Care*. Retrieved from Allianz Worldwide Care, International Medical Provider Finder:
<https://apps.allianzworldwidecare.com/poi/hospital-doctor-and-health-practitioner-finder?PROVTYPE=HOSPITALS&CON=Africa&COUNTRY=Kenya&CITY=Nairobi>

- Embrey, M. A. (1982). *MDS-3: MANAGING ACCESS TO MEDICINES AND HEALTH TECHNOLOGIES*. Kumarian Press.
- Emukule, G., Mott, J., Spreeuwenberg, P., Viboud, C., Commanday, A., Muthoka, P., . . . Paget, W. (2016). Influenza activity in Kenya, 2007-2013: Timing, association with climatic factors, and implications for vaccination campaigns. *Influenza and other respiratory diseases*, 375-385.
- Fattah, J., Ezzine, L., Aman, Z., Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International journal of Engineering Business Management*.
- Gupta, D. (2020). Periodic Inventory and Perpetual Inventory System. *Orderhive*.
- Gutierrez, R. S., Solis, A. O., & Mukhopadhyay, S. (2008). Lumpy demand forecasting using neural networks. *International Journal of Production Economics*, 409-420.
- Hurwitz, J., & Kirsch, D. (2018). *Machine Learning for dummies*. Hoboken: John Wiley & Sons, Inc.
- IBM. (2021). *Predictive Analytics*. Retrieved from IBM: <https://www.ibm.com/analytics/predictive-analytics>
- Ingersoll, K. (2015). *Inventory Management for the Pharmacy Technician*. Elite Professional Education.
- International Finance Corporation, UKAID, & IQVIA. (2020). *Kenya Pharmaceutical Industry Diagnostic Report*.
- Joseph, M., Joshin, J., & Kumar, S. (2013). New Trends in Healthcare Supply chain. *Academia*.
- Julianna Delua. (2021, March 12). *Supervised vs. Unsupervised Learning: What's the Difference?* Retrieved from IBM: <https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning>
- Kamalanabanand, G. E., & Premkumar, M. (2018). Medicine Box: Doctor's Prescription Recognition Using Deep Machine Learning. *International Conference on Advanced Trends in Computer Science & Information Technology*. 7. India: sciencepubco.

- Katie Ingersoll. (2017). *Inventory Management for the Pharmacy Technician*. ACPE.
- Keny, S., Nair, S., Nandi, S., & Khachane, D. (2021). Sales Prediction for Pharmaceutical Distribution Company. *International Journal of Engineering and Applied Physics(IJEAP)*, 186-191.
- Kerkkanen, A., Korpela, J., & Huiskonen, J. (2009). Demand Forecasting Errors in Industrial Context: Measurement and impacts. *International journal of Production Economics*, 43-48.
- Korn, S. (2011). The Opportunity of Predictive Analytics in Finance . *HPC Wire*.
- Kourentzes, N., Petropoulos, F., & Trapero, J. R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. *Int.J. Forecast*, 291-302.
- Laquanda, L., Kamal, A., & Peebles, D. (2017). Inventory Mangement Applications for Healthcare Supply Chains. *Researchgate*, 6.
- Linda, T., Joseph, C. M., & Ed, B. (2021). Predictive analytics vs. machine learning. *TechTarget*.
- Lu, Y., Chunxue, W., & Neal, X. N. (2022). An Intelligent Data Analysis System Combining ARIMA and LSTM for Persistent Organic Pollutants Concentration Prediction. *Electronics*, 11.
- Lucas, K. C., & James, N. L. (2010). Stock forecasting using Support Vector Machine. *2010 International Conference on Machine Learning and Cybernetics* (pp. 1607-1614). IEEE.
- Management science of health. (2012). MDS-3: Managing Access to Medicines and Health Technologies.
- Management Sciences of Health. (2012). *Ensuring Good Dispensing Practices*.
- Matsumoto, M., & Ikeda, A. (2015). Examination of demand forecasting by time series analysis for auto parts remanufacturing. *Springer*.

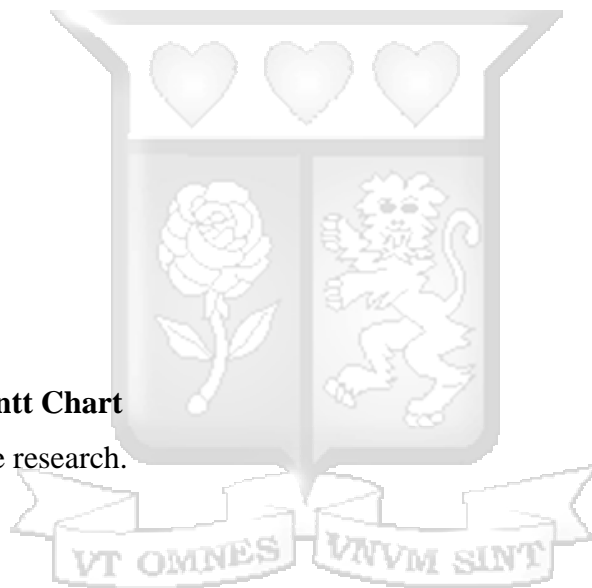
- Nicole, T., Clifford, L., Michele, B., & Kieran, S. (2014). Decentralized Automated Dispensing Devices: Systematic Review of Clinical and Economic Impacts in Hospitals. *The Canadian Journal of hospital pharmacy*, 138-148.
- Nursalam, Saafi, L. O., & Munir, S. (2020). The Relationship between planning and procurement of drugs and pharmaceutical services. *IOSR*, 24-28.
- Nysce, C. (2007). Predictive Analytics White Paper. *American Institute for Chartered Casualty Underwriters*, 1.
- Okoneko, T., Khrustsky, A. K., Veber, V. R., Egorova, E. S., & Antropova, G. A. (2019). *On The Optimization Of Public Procurement Of Medicines For Hospitals*. EpSBS.
- Peak Technologies. (2019, November 21). *RFID vs Barcode Comparison Advantages Diasadvantages*. Retrieved from Peak Technologies: <https://www.peaktech.com/blog/rfid-vs-barcode-comparison-advantages-disadvantages/>
- Petropoulos, F., Nikolopoulos, K., Spithourakis, G. P., & Assimakopoulos, V. (2013). Empirical heuristics for improving supply chain forecasting for intermittent demand SKUs, *Industrial Management. Data Syst*, 683-696.
- Pharmacy and Poisons Board. (2019, July). *Pharmacy and Poisons Board*. Retrieved from <https://www.pharmacyboardkenya.org/downloads>
- Sally, A. G. (2010). *Computational Learning Theory*. Washington University.
- Sammut, C., & Webb, G. I. (2010). *Encyclopedia of Machine Learning*. Boston , MA: Springer US.
- Shai, S.-S., & Ben-David, S. (2014). *Understandin Machine Learning from Theory to Algorithms*. New York: Cambridge University Press.
- Toroitich , A. M., Dunford, L., Armitage, R., & Tanna, S. (2022). Patients Access to Medicines – A Critical Review of the Healthcare System in Kenya. *Risk Management and Healthcare Policy*, 361-374.

- Trivedi, P., & Ashwani, S. (n.d.). Iterative waterfall. *A comparative study between iterative waterfall and incremental software development life cycle model for optimizing the resources using computer simulation.*
- Tsao, N. W., Lo, C., Babich, M., Bansback, N., & Shah, K. (2014). Decentralized automated dispensing devices: systematic review of clinical and economic impacts in hospitals. *The Canadian Journal of hospital pharmacy*, 67-138.
- Ukhalkar, P. (2018). Big Data Analytics Use Cases for Banking and Financial Services Industry. *Research Review Journals.*
- Wafula, F., Abuya, T., Amin, A., & Goodman, C. (2014). The policy-practice gap: describing discordances between regulation on paper and real-life practices among specialized drug shops in Kenya. *BMC Health Services Research.*
- Wang, C.-C., Chien, C.-H., & J.C. Trappey, A. (2021). Application of ARIMA and LSTM to predict Order Demand Based on Short Lead Time and On-Time Delivery Requirements. *MDPI.*
- WHO. (2015). Technical Consultation on Preventing and managing global stock outs of medicines. Geneva, Switzerland.
- World Bank Group. (2022, 07 31). *Climate Change Knowledge Portal.* Retrieved from <https://climateknowledgeportal.worldbank.org/country/kenya/climate-data-historical>

Appendices

Appendix A: Gantt Chart

Gantt chart for the research.



	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8
Identify research area								
Formulate research questions								
Formulate research strategy, research design and select methods								
Write research proposal								
Literature review								
Data collection								
Data analysis								
Design Model								
Validate model								
Model implementation								
Write final draft								
Thesis due								



Appendix B: Sample drug and weather data collected

Date	DrugName	Quantity
01/01/2011	Flugone Capsules	1
02/01/2011	Centriline 200	39
03/01/2011	Celestamine 300	34
04/01/2011	Desotor	23
05/01/2011	Monterllag	19
06/01/2011	Montifet	25
07/01/2011	Ascoril Range	27
08/01/2011	Rirathol	32
09/01/2011	Index	39
10/01/2011	Fungal dm	18
11/01/2011	Piriton syrup	23
12/01/2011	Centrizon syrup	4

1	name	datetime	tempmax	tempmin	temp	feelslikem	feelslikem	feelslike	dew	humidity	precip	precippro	precipcov	precip typesnow	snowdept	windgust	windspee	winddir
2	nairobi	01/01/2011	31.3	14.1	21.3	29.7	14.1	21	11.5	59.3	0	0	0				24.1	67.
3	nairobi	02/01/2011	30.3	14	20.6	28.5	14	20.5	11.1	60.4	0	0	0				33.5	51.
4	nairobi	03/01/2011	31.5	14	20.4	29.5	14	20.3	10.2	57	0	0	0				25.9	49.
5	nairobi	04/01/2011	27	14	20.8	26.5	14	20.7	10.6	55.2	0	0	0				24.1	67.
6	nairobi	05/01/2011	28.1	15	21	27.4	15	21	11.4	56.7	0	0	0				27.7	112.
7	nairobi	06/01/2011	28	16	21.7	27.1	16	21.6	11.5	55.4	1.3	100	54.17	rain			27.7	82.
8	nairobi	07/01/2011	32.7	1	21	30.7	-2	20.8	11.7	57.3	0	0	0				33.5	74.
9	nairobi	08/01/2011	27.1	15	20.7	26.5	15	20.6	9.1	50.8	0	0	0				29.5	73.
10	nairobi	09/01/2011	28	13	20.8	26.8	13	20.6	9	51.4	0.3	100	12.5	rain			22.3	125.
11	nairobi	10/01/2011	29	16	22.1	27.8	16	22	11.7	54.1	2.9	100	83.33	rain			24.1	13
12	nairobi	11/01/2011	29	16	22.3	27.9	16	22.1	11.9	54.9	3.4	100	70.83	rain			24.1	85.
13	nairobi	12/01/2011	27.1	18	22.1	26.7	18	22	13	58.1	2	100	66.67	rain			31.3	51.



Appendix C: Ethical Approval



25th March 2022

Ms Musanga Patience,
patience.musanga@strathmore.edu

Dear Ms Musanga,

RE: A Predictive Analytics Model for Pharmaceutical Inventory Management

This is to inform you that SU-IERC has reviewed and **approved** your above **SU masters'** research proposal. Your application reference number is **SU-IERC1225/21**. The approval period is **25th March 2022 to 24th March 2023**.

This approval is subject to compliance with the following requirements:

- i. Only approved documents including (informed consents, study instruments, MTA) will be used
- ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-IERC.
- iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-IERC within 48 hours of notification
- iv. Any changes, anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-IERC within 48 hours
- v. Clearance for export of biological specimens must be obtained from relevant institutions.
- vi. Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. Attach a comprehensive progress report to support the renewal.
- vii. Submission of an executive summary report within 90 days upon completion of the study to SU-IERC.

Prior to commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) <https://research-portal.nacosti.go.ke/> and obtain other clearances needed.

Yours sincerely,

for: **Dr Ben Ngoye,**
Secretary; SU-IERC
Cc: Prof Fred Were,
Chairperson; SU-IERC



Appendix D: Data Collection Approval



RELIAP

ARMACY
DABLE • QUALITY

armacy,
P.O.Box ,
Limuru Road,
Ru

RE: PERMISSION TO COLLECT DATA AT FARMACY.

This letter will serve as authorization for Musimbi Patience Musanga, a master's student from Strathmore University to collect data for her research project entitled, A Predictive Analytics Model for Pharmaceutical Inventory Management at Farmacy.

We acknowledge that we have reviewed the protocol presented, as well as the associated risks to the Facility and authorized the research data collection to proceed.

If we have any concerns or require additional information, we will contact the researcher.

Sincerely


Pharmacy Signatory
P. O. Box 62 - 00200,
NAIROBI

06/09/2022
Date

Appendix E: Sample code

```
import pandas as pd
import numpy as np
import plotly.express as xp
import pickle

import pandas as pd
import statsmodels.api as sm
from statsmodels.tsa.api import VAR
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.tsa.arima.model import ARIMA
from pandas.plotting import lag_plot
from statsmodels.tsa.stattools import adfuller
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

import matplotlib.pyplot as plt
import matplotlib
from sklearn.metrics import r2_score, mean_absolute_error
from sklearn.metrics import mean_squared_error as mse
from sklearn.metrics import mean_absolute_percentage_error as mape
import warnings
warnings.filterwarnings("ignore")

#read data
nairobi_df=pd.read_csv("nairobi.csv")
drug_df_train=pd.read_csv("Drug sales.csv")
```

```

nairobi_only.sort_values(by="datetime",inplace=True)
nairobi_df=nairobi_only[["datetime","temp","humidity","precip"]]
nairobi_df.drop_duplicates(subset=["datetime"],keep="first",inplace=True)
nairobi_df.reset_index(inplace=True,drop=True)

#drugs data
drug_df_train=drug_df_train[["DrugName","Date","Quantity"]]
drug_df_train["Date"]=pd.to_datetime(drug_df_train["Date"])
drug_df_train.sort_values(by="Date",inplace=True)
drug_df_train.sample(5)
#create main dataframe
main_df=pd.merge(drug_df_train,temp_df,left_on=["date"],right_on=["datetime"],how="left")

#create new columns
main_df["year"]=main_df["date"].dt.year
main_df["month"]=main_df["date"].dt.month
main_df["week"]=main_df["date"].dt.week

#sort and drop missing values
main_df.sort_values(by="date",inplace=True)
main_df.dropna(how="any",inplace=True)

#cumulative features
#get cumulative stats per week i.e total count of drugs per week and average temperature and humidity per month

#get cumulative stats per week i.e total count of drugs per week and average temperature and humidity per month
drug_sum=main_df.groupby(["DrugName","year","week"]).sum().reset_index().rename(columns={"Quantity":"total_weekly"})
humidity_temp=main_df.groupby(["year","week"]).mean().reset_index().rename(
    columns={"temp":"avg_wk_temp","humidity":"avg_wk_humidity","precip":"avg_wk_precip"})[["year","week","avg_wk_
main_df_agg2=pd.merge(drug_sum,humidity_temp,how="left",on=["year","week"])
#reset_index for each drug name
#reset_index for each drug name
main_df_agg2=main_df_agg2.groupby(["DrugName"]).apply(lambda x : x.reset_index(drop=True).drop(columns="DrugName"))

#dates for weeks
#lets get the start date for week..since we will be forecasting weekly values
dates_df=pd.DataFrame({"dates":pd.date_range(start='2011-01-01', end='2018-01-01',freq='1D')})
dates_df["year"]=dates_df["dates"].dt.year
dates_df["week"]=dates_df["dates"].dt.week
min_week_date=dates_df.groupby(["year","week"]).agg({"dates":"min"}).reset_index().rename(columns={"dates":"min_w
min_week_date
#merge with date
main_df_agg2=pd.merge(main_df_agg2,min_week_date,on=["year","week"],how="left")

clean_df=main_df_agg2[(main_df_agg2["year"]>=2014) & (main_df_agg2["min_week_date"]<pd.Timestamp(2017,10,1))]
clean_df.sort_values("min_week_date",inplace=True)

```

Python

Appendix F: Similarity index



Document information

Analyzed document	136507 B.docx (D141876344)
Submitted	7/8/2022 6:46:00 PM
Submitted by	
Submitter email	Patience.Musanga@strathmore.edu
Similarity	3%
Analysis address	library.strath@analysis.uikund.com

Sources included in the report

W	URL: https://analyticsindiamag.com/complete-guide-to-sarimax-in-python-for-time-series-modeling/ Fetched: 9/6/2021 6:25:34 PM		1
SA	28487314.pdf Document 28487314.pdf (D113101533)		1
W	URL: https://www.researchgate.net/publication/335674777_Autoregressive_Integrated_Moving_Average_ARIMA_Model_of_Forecast_Demand_in_Distribution_Centre Fetched: 1/8/2020 3:13:52 PM		1
SA	docmhemanthhh.docx Document docmhemanthhh.docx (D106040084)		1
SA	1920519_Nikunj Paliwal Final Draft.pdf Document 1920519_Nikunj Paliwal Final Draft.pdf (D136138400)		1
W	URL: https://towardsdatascience.com/time-series-forecasting-with-arima-sarima-and-sarimax-ee61099e78f6 Fetched: 5/18/2022 3:55:44 PM		1
W	URL: https://s3.amazonaws.com/EliteCME_WebSite_2013/f/pdf/RPTFL04IM14.pdf Fetched: 12/3/2019 1:19:47 PM		5
SA	Tech_Stack_Conference_Paper.pdf Document Tech_Stack_Conference_Paper.pdf (D116926331)		2
SA	RapportFeridFinalL.pdf Document RapportFeridFinalL.pdf (D117478978)		1

Entire Document

A Predictive Analytics Model for Pharmaceutical Inventory Management
Musimbi Patience Musanga 136507
A thesis submitted to the