Developing Predictive Analytics for ABC-VEN Matrix Inventory Management in Kenyan Hospitals

Lubanga Derrick Nyangule

169240

Submitted in partial fulfilment of the requirements for the degree of Master of Science in Data Science and Analytics of Strathmore University

Strathmore Institute of Mathematical Sciences
Strathmore University
Nairobi, Kenya

June 2025

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 dissertation contains no material previously published or written by another person except where due reference is made in the dissertation itself.

Student's Name: Lubanga Derrick Nyangule

Sign: **Date:** ____________________

Approval

The dissertation of Lubanga Derrick Nyangule was reviewed and approved for examination by the following:

Dr. Betsy Muriithi,
Lecturer, @iLab Africa,
Strathmore University

Dr. Godfrey Madigu,
Dean, Institute of Mathematical Sciences,
Strathmore University

Prof. Bernard Shibwabo,
Director of Graduate Studies,
Strathmore University

Abstract

Hospitals and healthcare facilities depend on robust inventory management systems to ensure that medications are available to patients promptly while keeping operational costs in check. Traditional inventory methods often fall short when it comes to adjusting for factors such as fluctuating demand, seasonal trends, and external disruptions. These shortcomings can lead to inefficiencies in determining restocking points and managing stock levels.

This study addresses these problems using the CRISP-ML(Q) methodology and in modeling, the power of machine learning (ML) algorithms and Mathematical optimization within the ABC-VEN matrix framework. The aim of the research is to accurately forecast the restock quantities and timing of pharmaceutical supplies by means of stochastic inventory optimization models based on Reorder Point (ROP) analysis. Past sales data, along with the ABC-VEN classification, was used to build a predictive model to estimate key inventory metrics. The hybrid approach then computed figures such as mean demand (μ_D), demand variability (σ_D), safety stock (SS), reorder point (ROP), ideal order size, stock levels, and total inventory cost. Finally, ROP forecasts were produced using these outcomes. The model's accuracy was validated using the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

With an R^2 score close to 0.85, the model proved to be highly effective, accounting for 85% of the variance in reorder point predictions. This performance underlines the model's capability in enhancing inventory control. Ultimately, the optimization model helped reduce both stockouts and overstocking by refining reorder point strategies and cutting down unnecessary expenses. This research highlights the valuable role that ML and optimization can play in strengthening hospital pharmacy operations while maintaining low costs and reliable drug availability.

KEY WORDS: Stockouts, Overstocking, EOQ, ROP, Machine Learning, Optimization, Model Performance

Table of Contents

List of Figures

List of Abbreviations

ABC-VEN	Always Better Control - Vital, Essential, Non-essential	AI	Artificial Intelligence
ANN	Artificial Neural Network	CRISP-ML(Q)	Cross-Industry Standard Process for Machine Learning (Quality)
DNN	Deep Neural Network	DRL	Deep Reinforcement Learning
EHR	Electronic Health Records	EOQ	Economic Order Quantity
HIS	Hospital Information System	KPI	Key Performance Indicator
LSTM	Long Short-Term Memory	MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error	ML	Machine Learning
MPC	Model Predictive Control	RMSE	Root Mean Square Error
RNN	Recurrent Neural Network	ROP	Reorder Point
RFID	Radio-Frequency Identification	SARIMAX	Seasonal ARIMA with Exogenous Variables
SDG	Sustainable Development Goals	SVM	Support Vector Machine

Acknowledgement

First and foremost, I extend my most sincere gratitude to God for His guidance, strength, and grace that have carried me throughout this journey. It is through His blessings that I have been able to finish this project. I am forever indebted to my project supervisor, Dr. Betsy Muriithi, for her directions, wisdom-filled counsel, and dedicated supervision along this journey. I also extend my heartfelt gratitude to Dr. Kennedy Senagi, Joseph Theuri, Dr John Olukuru and Dr Valentino Khoja for their encouragement and support during my research. Their expertise and enthusiasm have been instrumental in steering the direction and completion of my dissertation. My sincere appreciation to Strathmore University, and especially the Institute of Mathematical Sciences (SIMS) and the @iLabAfrica research facility, for their commitment to excellence in education and research. They have greatly impacted my academic and professional growth through the offering of world-class education and hands-on experience.

Dedication

I do this in honor of my parents, David Lubanga and Immaculate Lubanga, whose unwavering support, prayers, and motivation have been my ever-present pillar of strength. Your sacrifices, love, and confidence in me have been the foundation of my achievements. To my friends, siblings, and to all who have stood by me in this process, your comprehension and encouragement have meant the world to me. This achievement is equally yours as it is mine. Lastly, I dedicate this project to all the future researchers and students who aim to make a difference in the world through knowledge and determination. May this work inspire you to pursue your dreams with courage and determination.

Chapter 1

Introduction

1.1 Background to the Study

1.1.1 Introduction to Inventory management

Access to essential medicines continues to be a major challenge within Kenya's healthcare system. In 2018/2019, public health facilities reported an average availability of just 44%, compared to 72.4% in private institutions (Angula and Dongo, 2024). In Kenya's devolved county health system, Level 4–6 hospitals shoulder full responsibility for pharmaceutical inventory, yet they face frequent power outages, limited ICT infrastructure, and erratic transport links that introduce significant lead-time uncertainty and stock disruptions (Kirimi, 2023).This drugs results in shortage oftenly patients not receiving adequate care, leading to dissatisfaction and diminished trust in public healthcare services (Toroitich et al., 2022).

Hospital pharmacies are key players in healthcare delivery, responsible for ensuring timely access to medications and their safe distribution. To function efficiently and support the broader Hospital Supply Chain (HSC), robust inventory management practices are essential (Abu Zwaida et al., 2021a).

Inventory management refers to the systematic planning, organizing, and control of stock to strike a balance between minimizing inventory investment and meeting supply demands (Ali, 2011).It is essential for smooth operations and makes up a sizeable amount of working capital and assets in the majority of industries.

Beyond tracking stock levels, effective inventory management involves making strategic decisions in areas such as procurement, sourcing, financing, and marketing. These decisions are closely tied to organizational performance and carry financial risks if not handled properly (Ali, 2011). Given its broad impact across operational and financial domains, inventory management remains a rich field for research—offering the opportunity to improve business outcomes and reduce financial exposure. Among the eight essential roles of pharmacists outlined by the World Health Organization and the International Pharmaceutical Federation, resource management (encompassing money, materials, manpower, time, and information) stands out as a critical factor for professional success at both individual and organizational levels (Ali, 2011).In pharmaceutical practice, inventory refers to the stock of medications maintained to meet anticipated patient needs. It represents both the most substantial and the most liquid component of current assets within a pharmacy. The inventory value in the healthcare sector is growing due to the increasing cost and diversity of pharmaceuticals.

It is important to manage pharmacy inventory as it has direct impact on operational efficiency, financial performance and the ability to meet patient demands over time. However, as the pharmaceutical stock is dynamic and has been influenced by many factors, more work is needed to discover better strategies to cope with these complexities. Strategic work to improve inventory practices means there would be better financial oversight, fewer distractions for staff and better patient care. Inventory management changes as the power of big data, artificial intelligence (AI) and machine learning grows. Today pharmacists can make data driven choices with access to powerful tools that help them monitor stock levels better.

1.1.2 Predictive analytics in hospital pharmacy

Effective inventory management begins by clarifying your system's purpose: are you stocking raw materials to feed production, holding finished goods for direct customer orders, or managing inventory flow through distribution centers? (MSH, 2012). Predictive analytics has rapidly become a cornerstone in hospital pharmacy for anticipating medication needs and streamlining stock management. Musimbi(2022) developed a SARIMAX-based forecasting model that integrates historical sales with seasonal disease patterns to predict six weeks of drug requirements, achieving an RMSE of 5.5 and markedly reducing stockouts (Musimbi, 2022). Building on this, Ali-Nsour(2024) proposed a hybrid forecasting approach using Prophet for smooth demand, LSTM networks for erratic patterns, and Croston's method for intermittent demand within a Model Predictive Control framework, resulting in a 19% improvement in overall inventory efficiency through adaptive ordering (Al-Nsour, 2024). More broadly Adeghe (2024), reviewed the use of machine learning in pharmacy practice, highlighting its power to predict medication adherence and disease progression from EHR and wearable data, insights that can further refine demand forecasts and proactive stocking strategies (Adeghe et al., 2024).

1.1.3 Key Inventory Management Techniques

Pharmacies typically manage inventory using various methods, each with distinct strengths and limitations. Two foundational approaches include periodic and perpetual inventory systems. In periodic systems, stock levels are reviewed at regular intervals and replenished when they fall below predefined thresholds. Although simple to implement, this approach lacks real-time visibility and may fail to meet urgent demand. In contrast, perpetual inventory systems, more common in developed settings, use real-time digital tracking of stock movements, offering timely data but requiring significant technological investment.

While these systems provide useful context, the focus of this study is on the ABC and VEN matrix framework, a classification-based approach widely used in pharmaceutical inventory management. Rather than tracking every item in real time, ABC-VEN allows facilities to prioritize drugs based on financial value (ABC) and clinical importance (VEN), supporting cost-effective and risk-informed stock planning.

Pharmacies often complement these methods with safety stock to buffer against supply disruptions and fluctuating demand. Factors influencing safety stock include consumption patterns, supplier reliability, and lead time variability (Kumar et al., 2023).

Additionally, pharmacists must comply with regulatory standards for managing sensitive items like vaccines, biologicals, and controlled substances to ensure legal and ethical practices(Quick, 1982).

1.1.4 Importance of Inventory Management in Pharmacy

Effective inventory management contributes significantly to reducing both procurement and storage costs, thereby improving a pharmacy's overall financial performance. When inventory is managed well, it leads to healthier cash flow and enables funds to be redirected to other key operational areas (Ali, 2011).

Maintaining a consistent supply of essential medicines also prevents stockouts, which could otherwise jeopardize patient outcomes and erode confidence in the healthcare system (Ali, 2011). In addition, pharmacies are required to comply with stringent legal and professional standards for storing and handling pharmaceuticals. A strong inventory system supports this compliance and ultimately protects public health (Mawengkang et al., 2020).

1.1.5 Challenges in Hospital Pharmaceutical Inventory Management

Effective inventory control is essential for improving operational efficiency in pharmacies and healthcare institutions. However, several challenges complicate this process. One major hurdle is maintaining high-quality data, which is critical for accurate demand forecasting and stock management. Inaccurate or incomplete records can result in overstocking or shortages both of which can harm patient care and lead to financial strain (Ali, 2011).

The pharmaceutical supply chain is also inherently complex, involving multiple vendors, unpredictable lead times, and strict compliance requirements. This complexity necessitates sophisticated systems and workflows to ensure smooth inventory operations. Another pressing issue is the need to manage expiration dates carefully, as medicines have limited shelf lives. An efficient inventory system helps minimize waste by tracking expiry dates and reducing the likelihood of discarding unused stock.

To address these issues, pharmacies are increasingly turning to digital inventory systems, demand forecasting tools, and frameworks like ABC analysis. These tactics aid in achieving a balance between affordability and accessibility. Nevertheless, continuous assessment and modification are required to guarantee that inventory procedures continue to be in line with the dynamic healthcare environment.

1.2 Statement of the Problem

ABC-VEN is one of the most recognizable tools within the hospital pharmacies to pronounce the precise method of the pharmaceuticals' categorization depending on the cost and clinical importance. Despite helping with prioritization of inventory management as an approach, the framework has a major drawback of being more of a historical approach that categorizes items with little regard to future procurement needs. This absence makes it difficult for hospitals to maintain the right stock levels because the matrix does not indicate any reorder points or having to consider future needs of drugs.

To avoid such challenges, the necessity of developing a model to improve on the ABC-VEN matrix to foresee demand and set hi/lo flexible reorder points for drugs. Such a model would facilitate the objectives of health information systems so that future purchasing decisions can be made based on the availability of vital drugs and costs, reducing the possibilities of leaving stocks idle or expiring. Additionally, it would help hospitals in bringing alignment between inventory options and cash flow considerations, simultaneously improving operational functionality and financial planning.

Besides, this strategy aims at enhancing the ABC-VEN by employing the concept of predictive analytics to establish an effective inventory management system in responding to the needs of hospital pharmacists, and promoting enhanced healthcare delivery in Kenyan hospitals.

1.3 Research Objectives

1.3.1 General Objective

Developing an optimization model integrated with the ABC-VEN matrix for inventory management in Kenyan hospitals, enabling dynamic reorder level determination, optimizing stock levels, and aligning future drug purchases with organizational cash flow requirements.

1.3.2 Specific Objectives

1. Integrating the predictive model with hospital Health Information Systems (HIS) for automated inventory management..
2. To develop a streamlit application that will integrate the model and have main KPI points.

1.4 Research Questions

1. What are the key strengths and limitations associated with various existing predictive models for forecasting drug demand?
2. How can machine learning models improve reorder point prediction for ABC-VEN classified pharmaceuticals?

1.5 Significance of the Study

The proposed research is significant to hospital-based pharmacies because it offers knowledge on how to enhance their inventory management practices. Effective management of inventories is crucial for drug provision, cost optimization, and ultimately patient care improvement. By applying an optimization model, the research aims to present new approaches that have the potential to significantly enhance the effectiveness and precision of hospital pharmacy inventory management. One more advantage of the study will be cost saving and by improving the inventory systems, the study will be able to remove the excess stocks, and therefore preventing wastage of expiring drugs, which are costly outcomes in health facilities that are most frequently struggling with finance shortages. Besides, the study focuses on the significance of using the data in decision making based on the predictive models, the study offers hospital pharmacies practical solutions for improving demand forecasting. This predictive ability will help the hospital pharmacies to be in a position to order the stocks in advance so as to avoid cases of stock out or over stocking. In addition, this study has a significant positive impact on the United Nations Sustainable Development Goal of Good Health and Well-being (SDG-3) (Fernandez, 2020). Ultimately, the study contributes to the goal of healthy people and a healthy future for all ages through better management of pharmaceutical inventory. Inventory management is one of the critical success factors in any healthcare systems, and this study will assist in the provision of the relevant medicines to be accessed in the shortest time possible helping people stay healthy and maintain their health in the overall sense of the word.

1.6 Scope of the study

This study aims to develop an advanced pharmaceutical inventory optimization model that determines when to reorder and in what quantities while integrating real-time data for improved decision-making. The model seeks to minimize overall costs while ensuring adequate stock levels, addressing challenges related to stockouts and overstocking.

A data preprocessing and aggregation system will be structured to analyze product group demand patterns to form a robust inventory control system which adapts to changing consumption rates. Implementations of safety stock calculations for demand variability based on historical data and optimized reorder points (ROP). The system will also include supplier reliability and stock availability thresholds, together with lead time demand expected.

Stochastic inventory models were chosen to balance ordering, shortage and overstock costs within a single optimization framework, while our historical-data validation routines automatically trigger reorders only when meaningful demand shifts occur ensuring the system reflects real fluctuations and passes rigorous reliability checks.

All model assumptions and parameters were refined through a collaboration with one private Level 5 hospital, which signed an NDA to share a year of secondary usage and supply records and provided domain expertise in two structured workshops. Because we used existing transaction logs rather than surveying clinicians directly, there was no primary-data collection burden. Focusing on a single private facility also sidestepped many of the connectivity and procurement hurdles typical of public hospitals.

We deliberately opted for stochastic models over deep-learning methods because our dataset, while rich in time-stamped orders and stock movements, remains modest in size, and stochastic approaches offer transparent decision variables and lightweight computation, making them ideal for resource-constrained healthcare settings. Real-time feeds on supplier lead-time variability, ordering and storage costs, stock-out risk and seasonal demand patterns then power continuous, data-driven adjustments to reorder triggers and keep essential medicines available at minimal cost.

1.7 Justification

The research utilizes global and regional directives including **UNSDG Goal 3 (Good Health and Well-being)** as well as **UNSDG Goal 9 (Industry, Innovation, and Infrastructure)** and **Kenya's Vision 2030** of improving healthcare service delivery. By identifying potential drug demand using machine learning models it fosters positive change within the field of healthcare and aligns with **WHO** goal for affordable essential medication and **UNCTAD's** incorporated technology for enhancing healthcare systems resources. The findings of this research therefore present a solution by providing means to improve patient care, bring down costs and make healthcare systems more sustainable in the global arena.

Chapter 2

Literature Review

2.1 Introduction

A comprehensive literature review on pharmaceutical inventory management in hospital settings is presented in this chapter. The review is structured around (i) predictive modeling, (ii) data integration, and (iii) the application of emerging technologies, including artificial intelligence (AI) and machine learning (ML) (da Costa, 2023) It begins by looking at foundational studies on classical inventory approaches, as well as their limitations, and then presents recent efforts to enhance efficiency and accuracy. It aims to evaluate the current status of the field and develop innovative solutions to the common inventory challenges of hospital pharmacies.

2.2 Traditional Inventory Management in Pharmacies

Traditional inventory practices in pharmacies are often inefficient and financially burdensome, particularly when patients are unable to access prescribed medications. Understanding the shortcomings of these legacy systems is essential for driving meaningful improvements (da Costa, 2023). Conventional inventory control typically relies on three methods: visual inspection, periodic reviews, and perpetual tracking. While each method has its benefits, they also present significant drawbacks. Visual methods are simple but vulnerable to mistakes and oversight. Periodic checks are useful but may fail to account for sudden spikes or dips in demand. Perpetual systems offer better accuracy but demand continuous data input and require costly infrastructure investments (Ali, 2011).

Given the dynamic nature of healthcare shaped by outbreaks, seasonal patterns, and shifting patient needs, traditional inventory models are often insufficient. This has fueled growing interest in leveraging AI and ML technologies to modernize inventory practices and better respond to real-time demands.

2.2.1 Predicting drug shortages using pharmacy data and machine learning

Raman Pall's research highlights the serious consequences of drug shortages on patients, pharmacists, and the healthcare system at large. Using historical sales data from 22 Canadian pharmacies, the study applied machine learning techniques to anticipate future drug shortages. The models showed promising performance, predicting shortages in the most frequently dispensed drug groups with an accuracy of 69% and a kappa value of 0.44.

Several time series forecasting methods were explored, including Error Trend Seasonal (ETS) and ARIMA models. However, the best-performing approach was XGBoost, a gradient boosting algorithm known for its high accuracy, scalability, and speed. XGBoost builds an ensemble of weak learners typically decision trees, and incrementally improves prediction accuracy by correcting previous errors through boosting. This technique has become a leading choice for regression and classification tasks due to its strong performance and robustness.

The models were trained to use features such as drug availability per patient, historical shortage trends, duration of drug supply, and drug classification by therapeutic importance. This contributed to the enhancement of the model's capability to direct subsequent decision regarding inventory and effect of shortage on the cost. Further evidence that data driven prediction can aid pharmacy management is provided by the fact that the system also had the ability to forecast shortages a month in the future without input from direct supply chain data from the manufacturers. However, the models' performance, especially in predicting high-impact shortages, was limited, with accuracy dropping significantly for larger shortages. The assumption that manufacturers adhere strictly to the mandatory reporting protocol also brings into the picture potential gaps, as non compliance or delay in reporting might distorts the prediction. The need to incorporate additional points of supply chain data and enhance the model for high-value scarcity is felt (Pall et al., 2023).

2.2.2 Optimization of Inventory Management to Prevent Drug Shortages in the Hospital Supply Chain

Abu Zwaida et al. (2021b) looked at a supply chain model dealing with a practical problem of drug shortage by developing a framework called Dynamic Refilling Drug Optimization (DR2O). The purpose here was to reduce the cost of supply of products which encompass procurement costs, holding costs as well as the cost occasioned by stock out periods. The authors presented a solution adopting deep learning that involves decision making involving drug refills through Deep Reinforcement Learning (DRL) and Deep Neural Network (DNN) in a hospital. This model treats the drug inventory issue as a Markov Decision Process where each status of a medication is an MDP state and refilling decisions are made based on state status. Using simulations, they proved that their method would outcompete other existing methods of alleviating over-provisioning, ski-rental, and max-min approaches with regards to costs and stock shortages. Following this groundwork, the present study uses the DR2O model to address the problem of drug shortages. The DR2O framework seeks to reduce refilling costs at the same time as it address issues of storage cost and penalties for lack of stock. By means of Deep Reinforcement Learning (DRL) within the MDP setting, the model monitors inventory keeping the record of the inventory state of each drug as a state and makes refilling decisions based on rewards. Analysis of simulation results implied that this system should be considered as more effective than traditional approaches, as it established shorter shortages and overall expenses. However, the model main focus is on costs rather than on patients and clinical factors; for instance, the importance of some drugs with an extensive impact on patients' life (life-saving drugs) is not considered. These gaps point attention to the desirability of incorporating the criticality of medications in the model to ensure preferential stocking of crucial drugs. This type of an enhancement would turn the model to helpful more often in the health care facilities by relating cost effectiveness to usefulness, in helping to advance the patient care results.

2.2.3 Forecasting Drug Demand for Optimal Medical Inventory Management

Kumar et al. (2023) Kumar proposed a data-driven, integrated approach to predict the demand for drugs in health centers using state-of-the-art deep learning techniques like Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), and Ensemble Models.Their study showed the ability of these models in forecasting drug demand patterns based on past data to improve medical stock management to reduce stock out situations and excess stocks along the healthcare supply chain. The authors tested and compared several machine learning models, and found the Gradient Boost ensemble model as the best model, and it yields a MAPE as low as 1.98 % in the testing phase. The study fails to consider the essential nature of certain medications, particularly the prioritization of life-saving drugs within inventory management, even though it emphasizes the forecasting of drug demand trends. This shortcoming underscores the necessity for a more cohesive framework that aligns demand forecasting with clinical priorities to enhance healthcare outcomes. When incorporated with machine learning models, these factors become: might improve significantly the reliability of the drug demand forecasting. It is in the failure to integrate the broader external variables that research can develop a more accurate model for forecasting drug demand in hospital-based pharmacies by integrating a range of data sources.

2.2.4 Prescriptive analytics for inventory management in health care

Galli et al. (2021)suggested that drug replenishment was an appropriate field of prescriptive analytics because it could use stochastic optimization with machine learning 14 to solve the specific problems of hospital wards. Their study stressed that the use of drugs established itself in the hospital wards dependent on patient characteristics, for instance, the length of stay or severity of the disease, so demand is nonstationary. These scenarios were selected using machine learning methods such as Random Forest and XGBoost from the available wardlevel data and to make better prediction and decision regarding the drug replenishment. The model was designed to reduce the number of emergency orders and maintain low inventory levels using the historical data set and real-time features of wards. However, although, Galli's work enhances the hospital inventory management in the right manner by using the features of the problem, the work done by them is more of short-term and in a constrained hospital ward. They did not consider how the data from other categories like regional disruptions of drug supplies, seasonal diseases, or long-term trends may be incorporated to address the unique challenges in healthcare inventory management. The authors failed to explore methods which would integrate data from different categories including drug supply disruptions, seasonal diseases and long-term trends to solve specific inventory management challenges in healthcare. The authors stressed that patient-specific characteristics including length of stay and medical severity levels directly influence drug utilization rates in hospital wards thus making the demand patterns unpredictable. Ward-level data enabled the proposed approach to select scenarios through machine learning methods where Random Forest and XGBoost performed better for drug replenishment decision-making and prediction. The model reduced emergency orders and maintained low inventory levels through the combination of historical data and real-time ward features. However, while Galli's approach effectively improves hospital inventory management by incorporating relevant features, their work focuses on short-term predictions within a limited hospital ward environment. They did not explore the potential integration of broader data sources such as regional drug supply disruptions, seasonal disease patterns, or long-term predictions. This limitation has the potential to make research improve the models' prediction capabilities by incorporating various forms of external data that can be incorporated to widen the scope. better and more general forecasting solution for the drug demand in the hospital based environment.

2.2.5 Predicting medicine demand using deep learning techniques

Research by (Mousa and Al-Khateeb, 2023) highlights the importance of accurate demand forecasting in the improvement of the material supply chain in the pharmaceutical industry as well as minimizing the costs related to excess inventory. All these previous works utilized traditional time series models like ARIMA and linear regression models, which can then not capture much in terms of detail regarding the pattern of demand. To circumvent such limitations, the authors propose utilizing models of deep learning like RNN or LSTM networks, with which time series can be analyzed more effectively, and latent oscillations of demand can be identified.

The study employs different pharmacy data sets such as sales data set and prescription data set to train and test the proposed models. Random Forest (RF) and Support Vector Regression (SVR) are selected due to their appropriateness in managing big data and Long Short Term Memory (LSTM) networks are selected due to their capability to model and predict sequential data which is required to predict future demand based on past trends.

An important aspect of this study is the assessment of various algorithms for pharmaceutical forecasting. The paper shows that state-of-the-art models such as LSTM yield superior performance to baseline models in terms of the error rate when considering seasonality and non-linear demand patterns of the data.

However, this work does not include the extrinsic parameters such as, the current incidence of diseases, the prevailing climate or social, political and other conditions that play a very crucial role in the drug consumptions. However, the models largely based on historical sales and prescriptions fail to incorporate dynamic factors that could improve the responsiveness of the forecasts in a realistic healthcare environment.

2.2.6 ABC-VEN Pharmaceutical Inventory Management

Deressa applies the ABC-VEN matrix to improve pharmaceutical inventory by linking financial value (ABC) with clinical importance (VEN). ABC analysis categorizes items by cost impact Class A (high-value), B (moderate), and C (low-value) while VEN categorizes based on health significance Vital, Essential, and Non-essential. Combining both forms Category I (high-priority, high-cost or critical drugs), Category II (moderate priority), and Category III (low-cost, non-essential).

While effective for classifying and controlling inventory, the ABC-VEN method lacks predictive capabilities for forecasting future drug needs. This gap underscores the need to integrate predictive analytics into the framework to improve planning, budgeting, and stock availability (Magarsa Bayissa Deressa and Jemal, 2022).

2.2.7 Inventory Management in the Kenyan Healthcare Context

Inventory management in Kenyan healthcare systems, particularly within public and low-resource private facilities, faces critical challenges. Musimbi (2022) highlighted that Kenya's LMIS (Logistics Management Information System) suffers from limited integration across health facilities, inconsistent internet access, and manual data entry systems, which reduce the effectiveness of predictive tools and delay decision-making Musimbi (2022).

Karamshetty (2022) further emphasize that private healthcare facilities in Nairobi, despite being key providers of essential medicines, frequently encounter mismatches in drug stocking due to budget constraints, limited technical skills, and inadequate inventory systems Karamshetty et al. (2022). Facilities reported both understocking of essential medicines and overstocking of non-essential ones, largely due to a lack of reliable forecasting and human resource shortages.

These findings echo WHO's concerns that without digital infrastructure and well-trained personnel, even the most robust inventory models may fail to be adopted in LMICs. Therefore, solutions that incorporate lightweight machine learning algorithms and offline-compatible systems are more appropriate for Kenya's context.

2.2.8 Comparative Performance of Inventory Optimization Models

Table 2.1: Model Performance Comparison from Literature Review

Paper Title	**Authors & Year**	**Model/Method**	**Performance Metrics**	**Key Results**
Stochastic inventory management at a service facility with a set of reorder levels	Yadavalli et al. (2007)	Markov Arrival Process (MAP) with multiple reorder levels	Cost minimization with steady-state analysis	Local optimum achieved for service facility inventory; Hyperexponential arrival process tested
An Optimization Model for Hospitals Inventory Management in Pharmaceutical Supply Chain	Afnaria et al. (2020)	Multi-stage Stochastic Programming Model	**No numerical results provided**	Theoretical framework for handling demand, lead time, and received quantity uncertainties
Optimization of Inventory Management to Prevent Drug Shortages in the Hospital Supply Chain	Zwaida et al. (2021)	Deep Reinforcement Learning (DRLD)	**12.31% cost reduction** vs over-provisioning; **2.21% shortage rate** (best performance)	Best shortage reduction among all tested baseline methods; Convergence: 2800-4000 iterations
Optimizing Supply Chain Inventory: A Mixed Integer Linear Programming Approach	Vicente (2025)	Mixed Integer Linear Programming (MILP) with (s,S) policy	**0% optimality gap**; **128.80 seconds** computation time; **14,811 EUR** total cost	Guaranteed optimal solution for 15-period planning horizon
Stochastic models for inventory management at service facilities	Berman & Kim (1999)	Threshold Ordering Policy with Heuristic	**Within 2.5%** of optimal; **2/8 cases** achieved optimal cost	Near-optimal performance with simple heuristic; Better at higher server utilization

2.3 Challenges in Implementing AI and ML

Despite their promise, the adoption of AI and ML in pharmaceutical inventory management faces several barriers. Integration with existing pharmacy systems is difficult, especially since many are built on traditional infrastructure and require substantial financial and human resource investments (Angula and Dongo, 2024). Data quality poses another major hurdle , AI models depend on large, clean datasets, yet many hospital pharmacies struggle with data entry errors, missing records, and inconsistent formats (Ali, 2011; Angula and Dongo, 2024).

Additionally, the complexity of AI algorithms often leads to mistrust and resistance from healthcare staff unfamiliar with such technologies. Ethical concerns and privacy risks also present major obstacles, as these systems handle sensitive patient data that must comply with strict privacy regulations (Kumar et al., 2023). Finally, sustaining AI models is costly, requiring ongoing updates, monitoring, and skilled personnel to adapt to evolving drug demand and usage patterns.

2.4 Research Gaps

Current processes associated with stock management in hospital pharmacies use conventional methods consisting of classification, the most well-known of which is the ABC-VEN analysis. Despite the success of the ABC-VEN framework in inventory prioritization, the technique does not look into the future or assist in fluctuating reordering margins. Current applications do not support predictive analytics, and so hospitals cannot adjust inventory use or correlate drug buying to financial enhancing. In addition, most explicit models are mainly designed by minimizing costs and do not necessarily consider clinical necessities such as the fact that some drugs are more life-saving than others. While current day big data and machine learning techniques have had a significant success in demand forecasting, existing solutions do not have operational integration with common inventory management techniques like the ABC-VEN and do not offer handles for operational inventory adjustments in real time mode. To fill these gaps, we propose to build a new quantitative model to expand the ABC-VEN system, integrating the relative priorities of clinical fields and adapting reorder factors about inventory and financial management both in healthcare organizations.

Chapter 3

Methodology

3.1 Introduction

This chapter outlines the methodological framework adopted for the study, which focuses on applying optimization models to improve pharmaceutical inventory management in hospitals. It presents the chosen research design, methods of data collection, and analytical techniques used to address the study's objectives.

The section begins by detailing the rationale for the selected research design, referencing prior work such as that by Schröer et al. (2021). It then introduces the data sources and tools employed to gather relevant information, followed by an explanation of the analytical processes used to evaluate the effectiveness of predictive models in forecasting drug demand and optimizing stock levels.

3.1.1 Research Design

Research Approach

To guide the research process in a structured way, the study followed the CRISP-ML(Q) methodology,the Cross-Industry Standard Process for Machine Learning enhanced with quality assurance measures. This framework provides a structured, traceable lifecycle for machine learning development, particularly suited for healthcare contexts where model transparency, governance, and ongoing monitoring are critical (Studer et al., 2021).

Compared to alternatives such as TDSP, which is tailored for Microsoft ecosystems, or KDD, which emphasizes data mining without robust post-deployment strategies, CRISP-ML(Q) integrates continuous validation and stakeholder feedback loops. Its focus on embedding quality controls at every stage ensures the development of reliable, explainable, and sustainable ML systems for health applications.

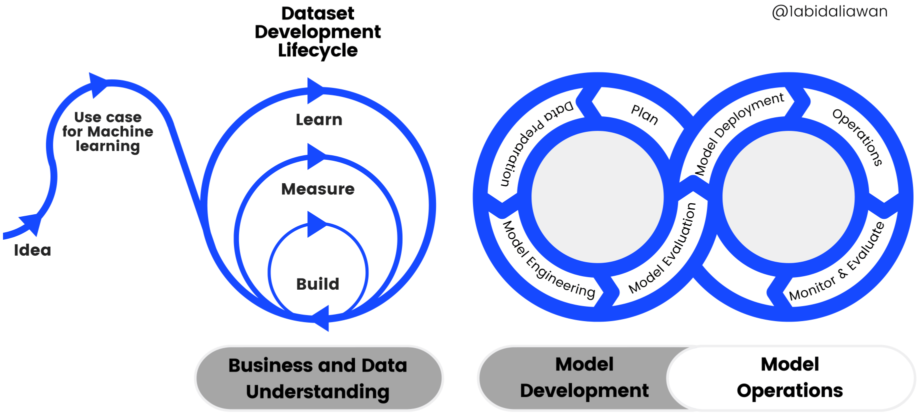
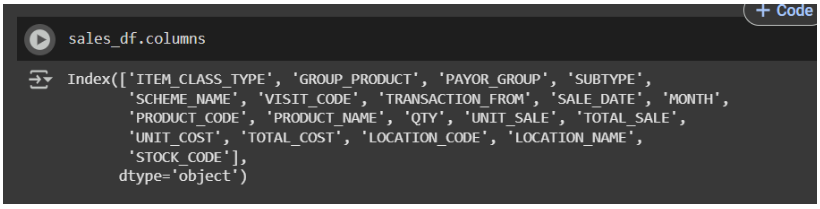


Figure 3.1: CRISP-ML(Q)
Source: KDnuggets (2022) – Making Sense of CRISP-ML(Q)

3.1.2 Business Understanding

The primary objective was to optimize drug inventory management in hospitals by employing mathematical models for reorder levels and inventory control. These models will ensure accurate orders for future drug demand, minimizing stockouts and overstocking while reducing costs related to emergency procurement, drug wastage due to expiration, and high storage fees. By integrating historical sales data with Holding stock costs, shortage costs, and overstocking costs, the models provide precise and data-driven insights, enabling efficient pharmaceutical logistics and resource allocation (Mawengkang et al., 2020).



```
sales_df.columns
```

```
Index(['ITEM_CLASS_TYPE', 'GROUP_PRODUCT', 'PAYOR_GROUP', 'SUBTYPE',
       'SCHEME_NAME', 'VISIT_CODE', 'TRANSACTION_FROM', 'SALE_DATE', 'MONTH',
       'PRODUCT_CODE', 'PRODUCT_NAME', 'QTY', 'UNIT_SALE', 'TOTAL_SALE',
       'UNIT_COST', 'TOTAL_COST', 'LOCATION_CODE', 'LOCATION_NAME',
       'STOCK_CODE'],
      dtype='object')
```

Figure 3.2: column overview of the sales data

3.2 Data Understanding

This study utilized inventory records from a private hospital covering the period between January 2023 and December 2024. The data, reported as secondary sales information, was aggregated at the sales level, meaning it did not include any patient identifiable information. Monthly Excel spreadsheets were consolidated into a single dataset and analyzed using the Python library `pandas`.

Since no patient-level data was used, there were no direct privacy concerns. Nevertheless, the study adhered to ethical research practices and aligned with Kenya's Data Protection Act (2019) to ensure responsible handling of health-related data.

The main variables in the final data set were the date of sale, the name of the product, the classification group, the unit cost, the unit sales, and the quantity sold. Each item is classified using the ABC-VEN model represented by variables `GROUP_PRODUCT` (Deressa et al., 2022).

While the data was rich and well-structured, the reliance on a single hospital limits the model's generalizability, particularly to institutions not applying the ABC-VEN framework. This limitation is acknowledged and addressed as a key area for improvement through multi-institutional validation in future work.

Each entry in the dataset presents individual sales data regarding units of products analyzed for time-based demand patterns. The long duration of the dataset enabled thorough analysis of standard drug consumption behavior. The ABC-VEN classification method (Mohammed and Workneh, 2020) creates an organized system which helps strategic inventory management to minimize costs while maximizing efficiency.

3.3 Data Preparation

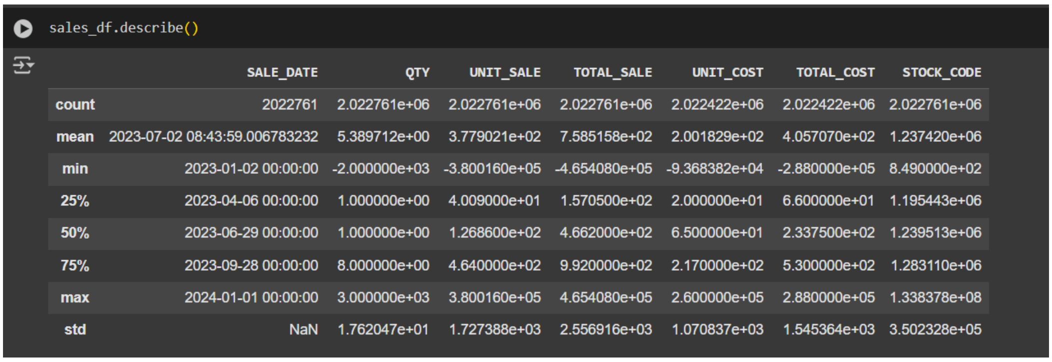
This dataset has 2,022,761 rows on 19 columns. The dataset includes product related information as well as sales transactions data along with quantity and cost details and supporting elements such as product codes and location specifications. Optimization modeling was used to solve the problem with the preprocessed data from a dataset that had to go through several steps to achieve the desired result. Data cleaning was conducted before starting. The data that was collected was processed in the preprocessing stage by engineering important features and building a structured data structure.

3.3.1 Data Cleaning

In the initial step, the raw dataset was thoroughly examined to identify inconsistencies, primarily caused by missing values and duplicate entries. Particular attention was given to critical columns such as `QTY`, `PRODUCT_NAME`, and `SALE_DATE`, where missing values were carefully inspected. Records with minimal missing data were removed, while those with more substantial gaps were subjected to imputation using techniques like historical trend imputation.

To ensure consistency in the `SALE_DATE` column, date standardization procedures were applied. Additionally, new temporal features such as `MONTH` and `YEAR` were derived, and columns like unit price and quantity sold were converted from string to numeric data types. These transformations ensured the data was correctly formatted for numerical analysis. Duplicate records were also removed to reduce redundancy and maintain the integrity of the dataset.

Several medicines had missing sales records an artifact of a past overstocking event in the private hospital. We first imputed those gaps using the median weekly sales for each ABC-VEN category. To address the resulting class imbalance, we then applied SMOTE (Synthetic Minority Over-sampling Technique) to boost under-represented medicines in the training set. (We'll flag both the imputation and synthetic sampling as limitations in Chapter 7.)



```
sales_df.describe()
```

```
                          SALE_DATE            QTY      UNIT_SALE     TOTAL_SALE      UNIT_COST     TOTAL_COST     STOCK_CODE
count                       2022761   2.022761e+06   2.022761e+06   2.022761e+06   2.022422e+06   2.022422e+06   2.022761e+06
mean  2023-07-02 08:43:59.006783232   5.389712e+00   3.779021e+02   7.585158e+02   2.001829e+02   4.057070e+02   1.237420e+06
min             2023-01-02 00:00:00  -2.000000e+03  -3.800160e+05  -4.654080e+05  -9.368382e+04  -2.880000e+05   8.490000e+02
25%             2023-04-06 00:00:00   1.000000e+00   4.009000e+01   1.570500e+02   2.000000e+01   6.600000e+01   1.195443e+06
50%             2023-06-29 00:00:00   1.000000e+00   1.268600e+02   4.662000e+02   6.500000e+01   2.337500e+02   1.239513e+06
75%             2023-09-28 00:00:00   8.000000e+00   4.640000e+02   9.920000e+02   2.170000e+02   5.300000e+02   1.283110e+06
max             2024-01-01 00:00:00   3.000000e+03   3.800160e+05   4.654080e+05   2.600000e+05   2.880000e+05   1.338378e+08
std                             NaN   1.762047e+01   1.727388e+03   2.556916e+03   1.070837e+03   1.545364e+03   3.502328e+05
```

Figure 3.3: Descriptive statistics of the data

float

3.3.2 Exploratory Data Analysis

We performed exploratory data analysis (EDA) to examine sales patterns and product performance and key variables in this step. CE products were the most common category within `GROUP_PRODUCT` while BE products ranked second which demonstrates consumables make up most hospital sales.

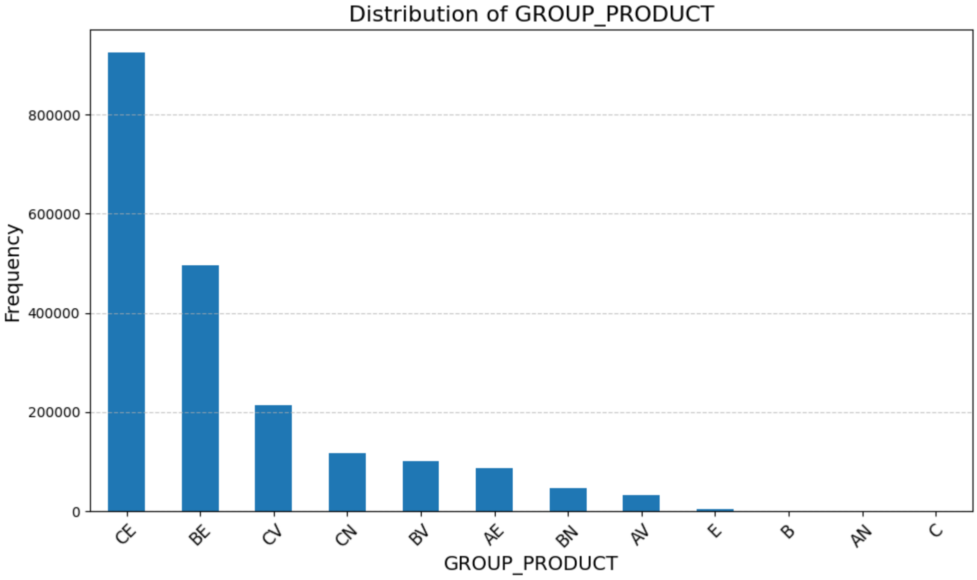


Figure 3.4: Product Group Distribution

The analyzed pharmaceutical product groups displayed diverse sales trends which changed across various product categories. The analyzed products showed either steady predictable performance or highly variable seasonal fluctuations.

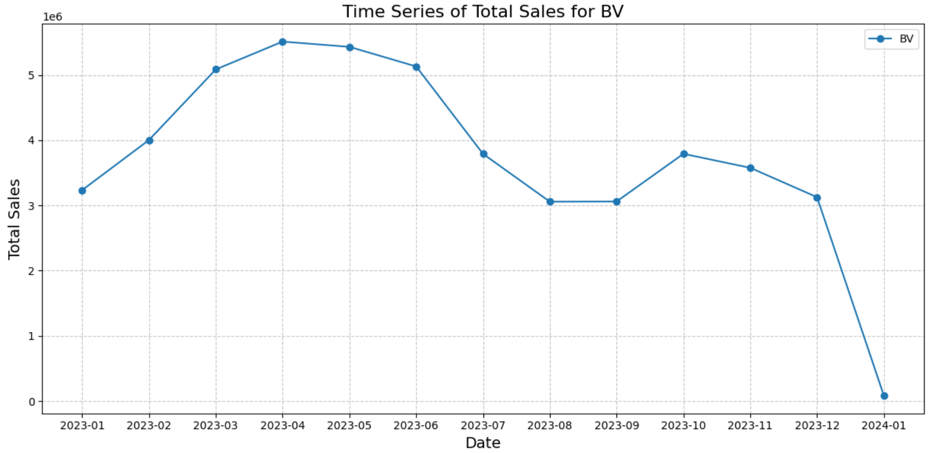
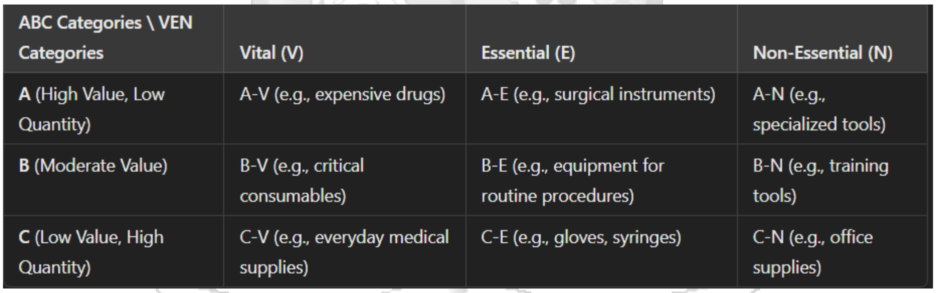


Figure 3.5: Line graph for product group BV



ABC Categories \ VEN Categories	Vital (V)	Essential (E)	Non-Essential (N)
A (High Value, Low Quantity)	A-V (e.g., expensive drugs)	A-E (e.g., surgical instruments)	A-N (e.g., specialized tools)
B (Moderate Value)	B-V (e.g., critical consumables)	B-E (e.g., equipment for routine procedures)	B-N (e.g., training tools)
C (Low Value, High Quantity)	C-V (e.g., everyday medical supplies)	C-E (e.g., gloves, syringes)	C-N (e.g., office supplies)

Figure 3.6: ABC-VEN matrix classification

3.3.3 Feature Engineering

Several important features were developed to improve the inventory optimization dataset. The sales quantities were aggregated monthly through `GROUP_PRODUCT`, `PRODUCT_NAME`, and `MONTH` to produce structured demand time series data. The analysis calculated mean demand (μ_D) and standard deviation of demand (σ_D) for each product to determine its demand variability, as this factor significantly impacts stock optimization.

The data set includes lead time data by assuming constant three-day lead time, which is necessary to maintain adequate inventory levels. From this data, the Reorder Point (ROP) and Safety Stock (SS) were determined for each product. The safety stock was calculated using the typical formula:

$$SS = Z \cdot \sigma_D \cdot \sqrt{\text{Lead Time}} \tag{3.1}$$

where Z represents the service level factor, set at 95% confidence. The reorder point was calculated as:

$$ROP = (\mu_D \times \text{Lead Time}) + SS \tag{3.2}$$

These calculations help ensure that stock levels are sufficient to meet demand fluctuations while minimizing excess inventory (Ebojoh et al., 2021).

3.3.4 Data Transformation

We began by calculating the **mean demand** (μ_D) for each product using historical sales data. Next, we computed the **standard deviation of demand** (σ_D) to understand demand variability. We estimated the **lead time** (L) as 3 days, based on information provided by the hospital.

Hospital records were utilized to compute the values of holding cost (H) and ordering cost (S). Additionally, the assessment involved determining the shortage cost (C_s) and overstock cost (C_o) to address instances of stockouts and surplus inventory. These numbers were used to determine the safety stock (SS), which is a buffer for both demand variability and lead-time variability. The safety stock that was calculated was added to the average lead-time demand in order to determine the Reorder Point (ROP), allowing for more reliable and economic inventory management.

The data required several adjustments to align with the specifications of the optimization model. Demand values were normalized to eliminate scale-related biases across different products. The `GROUP_PRODUCT` feature was encoded categorically to enhance processing efficiency during the implementation of the optimization model. This preprocessing sequence enabled the construction of essential variables necessary for formulating the optimization model, which was designed with cost minimization as its primary objective.

3.3.5 Final Prepared Dataset

The structured dataset was ready for modeling and analysis following all preprocessing operations. The created dataset contained essential information about sales dates along with product names and groups and demand quantities yet included engineered features for demand variability and reorder points and safety stock calculations. The structured dataset functioned as the foundation for the optimization model to run mathematical calculations that handled inventory management effectively.

3.3.6 Modeling

Pharmaceutical inventory management is not efficient in hospitals and thus this research develops an optimization model to minimize total inventory cost while ensuring proper stock levels for patient demand. The model follows important cost drivers which guide inventory decisions by considering holding costs and ordering costs and shortage costs and overstocking costs.

The total inventory cost (TC) is given by the following equation:

$$TC = h \cdot I \cdot T + \frac{K}{Q} + s \cdot \max(0, ROP - \mu_D \cdot LT - I) + o \cdot \max(0, I - ROP) \tag{3.3}$$

Where:

(i) h: Holding cost/unit in per

(ii) I: Average inventory level (units held)

(iii) T: Time period (days)

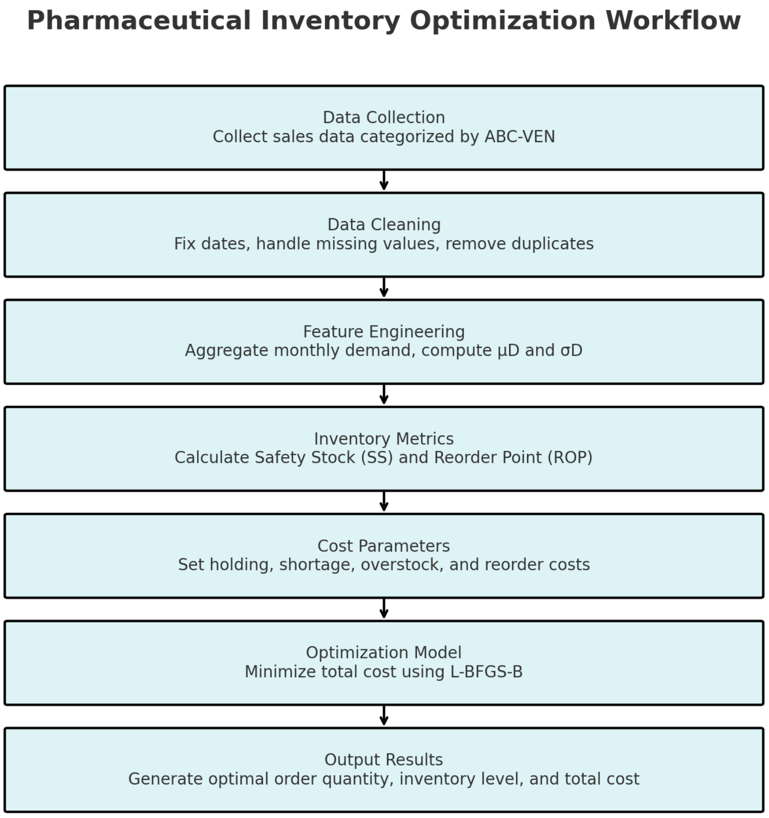


Figure 3.7: Workflow diagram illustrating the step-by-step process of pharmaceutical inventory optimization.

(iv) K: Fixed ordering cost per order

(v) Q: Order quantity

(vi) s: Shortage cost per unit

(vii) ROP: Reorder points

(viii) μ_D: Mean daily demand

(ix) LT: Lead period in days

(x) o: Overstock cost per unit

3.3.7 Cost Component Breakdown

1. Holding Cost ($h \cdot I \cdot T$)

This term represents the cost of storing inventory over a specified period. It increases with the number of units stored and the duration of storage. Properly estimating h is critical as it incorporates warehousing, insurance, and other indirect costs.

2. Ordering Cost ($\frac{K}{Q}$)

This encompasses the ordering cost that is incurred on every purchase order. If order size Q is small, many orders are produced, and as a result, cumulative ordering cost is high. A large Q reduces the purchases but increases the holding costs.

3. Shortage Cost ($s \cdot \max(0, ROP - \mu_D \cdot LT - I)$)

This term penalizes stockouts. If the current inventory I is less than the expected demand during lead time, a shortage occurs. The `max` function ensures that costs are only calculated when shortages happen.

4. Overstock Cost ($o \cdot \max(0, I - ROP)$)

This reflects the cost of excess inventory beyond the reorder point. It accounts for risks like expiry, wastage, and excess capital tied up in inventory.

3.3.8 Supporting Calculations

Reorder Point (ROP) is determined using expected demand and a buffer called Safety Stock:

$$ROP = \mu_D \cdot LT + SS \tag{3.4}$$

Safety Stock (SS) accounts for demand variability:

$$SS = Z \cdot \sigma_D \cdot \sqrt{LT} \tag{3.5}$$

Where:

(i) Z: Z-score for desired service level (e.g., 1.645 for 95%)

(ii) σ_D: Standard deviation of demand

3.3.9 Decision Variables

The primary decision variables optimized in this model are:

(i) **Order Quantity (Q)**: How much stock to order when replenishment is triggered.

(ii) **Inventory Level (I)**: The average level of stock to be maintained.

The L-BFGS-B algorithm and other numerical methods optimize variables to achieve minimum total cost based on constraints from historical sales data input parameters. The optimization model functions as the base for hospital pharmacy inventory decision guidance and reorder point prediction.

3.3.10 Optmization Model output

High ROP values in **BE** and **CE** product categories indicate either high customer demand or significant demand variability, necessitating increased stock levels and continuous monitoring

	GROUP_PRODUCT	PRODUCT_NAME	µD	σD	SS	ROP	STR	Optimal Order Quantity	Optimal Inventory Level	Total Cost
133	BE	GLEMONT-L.7S (TAKE AT NIGHT)	1912.000000	1769.181167	5040.344162	10776.344162	2.0	5736.000000	5040.344162	5040.352879
66	BE	CIDEX OPA 5LT	27518.153846	11196.214835	31897.680772	114452.142310	13.0	82554.461538	31897.680773	4907.336109
68	BE	CIPLADONE 1G TABLETS(DISSOLVED IN 1/2 GLASS W...	21089.153846	7345.855043	20928.121029	84195.582568	13.0	63267.461538	20928.121031	3219.711718
4	AE	LOSARTAS-HT TABS.	8572.000000	2609.686699	7434.919253	33150.919253	13.0	25716.000000	12857.846149	2546.695165
138	BE	GOFEN SOFTGEL 400MG CAPSULES 10S	1553.000000	1297.380438	3696.198014	8355.198014	3.0	4659.000000	3696.198014	2464.142741
308	BN	MEDI-KEEL LOZENGES (HON&LEM)	4841.777778	3759.504747	10710.716439	25236.049772	9.0	14525.333333	10710.716439	2380.162651
254	BE	STERANIOS 2% NG SOLUTION	1666.666667	1154.700538	3289.707254	8289.707254	3.0	5000.039753	3289.707255	2193.148170
5	AE	MONTEL (LEVOCETRIZINE/MONTELUKAST) (TAKE AT NI...	6734.538462	2251.857626	6415.474941	26619.090325	13.0	20203.615385	10101.653841	2112.318947
293	BE	XYKAA EXTEND 1000MG (PARACETAMOL) TABS	12034.307692	4050.294586	11539.167980	47642.091057	13.0	36102.923077	18051.307710	1798.001555
1933	CE	PANADOL ADVANCE TABS 500MG	8287.230769	3998.191115	11390.726752	36252.419060	13.0	24861.692308	12430.692325	1769.018312

Figure 3.8: Model output

to ensure product availability during lead time periods. In contrast, ROP values in **B** and **E** categories suggest lower demand or minimal variability, resulting in less frequent ordering and smaller inventory quantities. Consequently, the total inventory costs for **BE** and **CE** products tend to be higher due to the need for larger order quantities to meet demand. On the other hand, **E** and **B** products incur lower total costs, reflecting their smaller order sizes and consistent, low-level demand.

3.4 Machine Learning Implementation

To enhance decision-making in inventory management, we employed machine learning techniques to predict reorder points (ROP) by processing key features such as average demand (μ_D), demand variability (σ_D), and stock turnover rate (STR). The modeling approach was carefully designed to minimize overfitting and optimize predictive accuracy, utilizing grid search for hyperparameter tuning and cross-validation to ensure generalizability to unseen data. aining accurate prediction capability.

3.4.1 Feature Selection

All candidate predictors were derived from our optimization model mean demand (μ_D), demand variability (σ_D), lead time, unit cost, and ABC–VEN category and were first evaluated via Pearson correlation analysis against the reorder-point (ROP), retaining only those with an absolute correlation coefficient $|\rho| > 0.3$. We then applied Lasso regression to rank feature importance, preserving only predictors with non-zero coefficients.

3.4.2 Data Splitting

The modeling process began with splitting the dataset into training and test subsets. This was achieved by using the `train_test_split` function of the `sklearn.model_selection` module. The dataset was divided into two sets according to the process below:

(i) **80% Training Set** – Used to train the model and learn patterns.

(ii) **20% Testing Set** – Used to evaluate the model's performance on unseen data.

This approach ensures that the model is exposed to most of the available data during training while maintaining a separate portion to validate its generalization ability on new, unseen examples.

3.4.3 Data Standardization and encoding

In this step, the raw data was transformed into a format suitable for machine learning models. A `ColumnTransformer` was implemented to handle both numerical and categorical features. Numerical features were standardized using the `StandardScaler`, which removes the mean and scales the data to have unit variance. This standardization ensures that the learning algorithms treat all numerical features equally, preventing any one feature from dominating due to scale differences.

Categorical features, such as the `GROUP_PRODUCT` variable, were encoded using `OneHotEncoder`. This transformation converts categorical values into binary indicator variables, enabling models to understand categorical distinctions without introducing ordinal relationships where none exist.

3.5 Model Selection Summary

We selected three models for predicting reorder points (ROP): **Random Forest Regressor (RF)**, **Gradient Boosting Regressor (GB)**, and **Linear Regression**. We selected RF and GB because these algorithms excel at processing non-linear data relationships through ensemble methods where RF averages tree predictions and GB uses sequential trees to fix errors. We used Linear Regression as the basic comparison point. The chosen models served both performance efficiency and complexity balance to provide accurate inventory prediction capabilities.

3.5.1 Bias-Variance Tradeoff and Model Selection

Intrinsic regularization factors and complexity adjustment assisted in reducing overfitting for the models. Random Forest (RF) and Gradient Boosting (GB) were the options because they were very competent for regression tasks. Stacking ensemble was applied to blend the best performing factors of the models. The main emphasis was on having a balance between variance and bias to provide a good generalization for new cases.

3.5.2 Modeling Strategy and Configuration

Random Forest (RF): RF was constrained by limiting tree depth and requiring a minimum number of samples per leaf. This forced the model to focus on broad patterns rather than memorizing noise. Parameters such as `max_depth=3-5`, `min_samples_leaf=30-50`, and a limited number of trees were applied.

Gradient Boosting (GB): GB was trained on several shallow trees and a small learning rate. Early stopping was employed to prevent training when validation performance stopped improving. Shallow tree depth (e.g., 2), subsampling, and minimum leaf size were also utilized to lower model complexity.

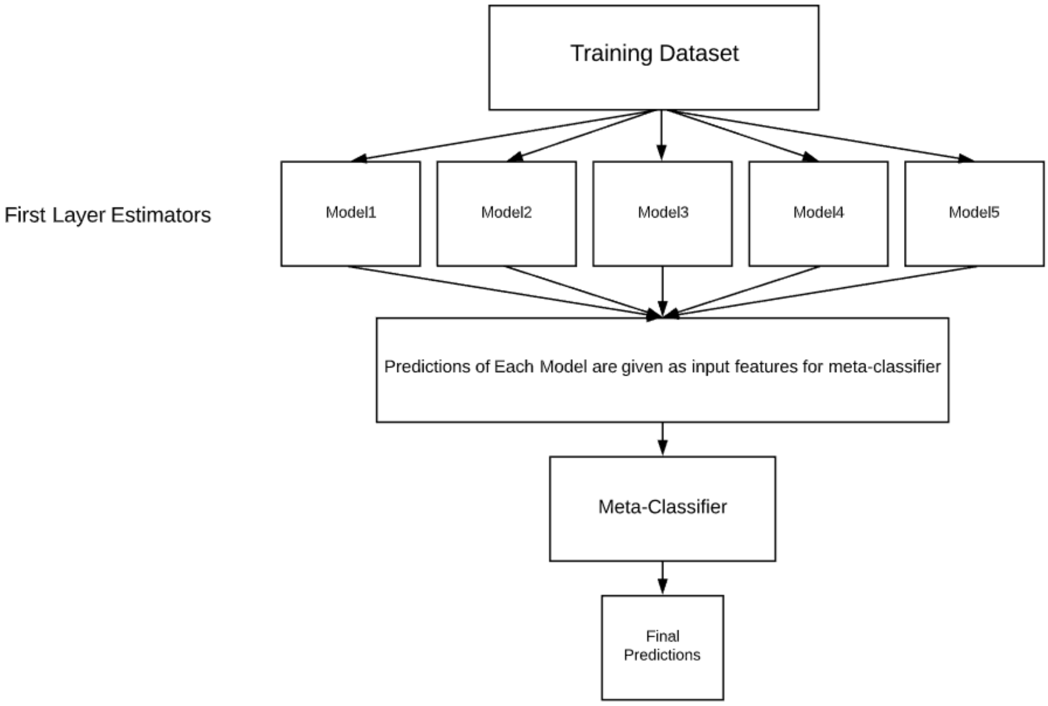


Figure 3.9: Stacking ensemble architecture

Stacking Regressor: Predictions from RF and GB were fed into a meta-model a simple linear regressor which learned to combine their outputs. Cross-validation was used to generate the out-of-fold predictions for meta-training, avoiding data leakage.

Each model approximated the following mapping:

$$ROP = f(\mu_D, \sigma_D, STR, GROUP\_PRODUCT) \tag{3.6}$$

The stacked model was trained as:

$$ROP_{stacked} = \beta_1 ROP_{RF} + \beta_2 ROP_{GB} + \varepsilon \tag{3.7}$$

This careful configuration enabled the development of a predictive model that supports operational planning while maintaining reliability across datasets (Ahrens et al., 2022).

3.6 Evaluation of the Models

The performance of the predictive models was assessed using widely accepted regression evaluation metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). These indicators provided insight into how closely the model's predicted reorder points aligned with actual outcomes in the test dataset. The model achieving the strongest results across these metrics was chosen for final deployment (Emmert-Streib and Dehmer, 2019).

3.6.1 Root Mean Squared Error (RMSE)

RMSE measures the square root of the average of the squared differences between actual and predicted values. It is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.8}$$

Where:

(i) y_i: Actual ROP value

(ii) $\hat{y}_i$: Predicted ROP

(iii) n: Number of observations

RMSE penalizes larger errors more heavily due to the squaring, making it particularly useful in applications like inventory management where large deviations (e.g., significant underestimation or overestimation of reorder points) Hodson (2022) can lead to costly stockouts or overstocking.

3.6.2 Mean Absolute Error (MAE)

MAE captures the average of the absolute differences between predicted and actual values:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.9}$$

Unlike RMSE, MAE treats all errors equally without emphasizing outliers. It provides a straightforward interpretation of the average error made by the model in predicting the reorder point (Willmott and Matsuura, 2005). In the hospital inventory context, it offers an estimate of the typical number of units by which the model's prediction deviates from the true reorder requirement.

3.6.3 Coefficient of Determination (R^2)

R^2 is the proportion of variance in the dependent variable that can be accounted for by the independent variables:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{3.10}$$

Where $\bar{y}$ is the mean of the actual ROP values.

An R^2 value close to 1 indicates that the model explains most of the variability in the data, while a value closer to 0 indicates poor explanatory power (Di Bucchianico, 2008).The project maintained R^2 values below 0.85 to avoid overfitting during the model development process. By constraining the coefficient of determination to values less than or equal to 0.85, the model preserved its robustness and improved generalization, making it more reliable for deployment on real-world data.

3.6.4 Model Performance Results

The performance assessment of trained models took place on the test dataset. The Stacking Regressor achieved superior outcomes than single models by delivering optimal accuracy together with generalizability..

Model	**RMSE**	**MAE**	R^2
Random Forest Regressor	105	22	0.80
Gradient Boosting Regressor	108	28	0.82
Stacking Regressor	103	20	0.84

Table 3.1: Model Performance Results on Test Data

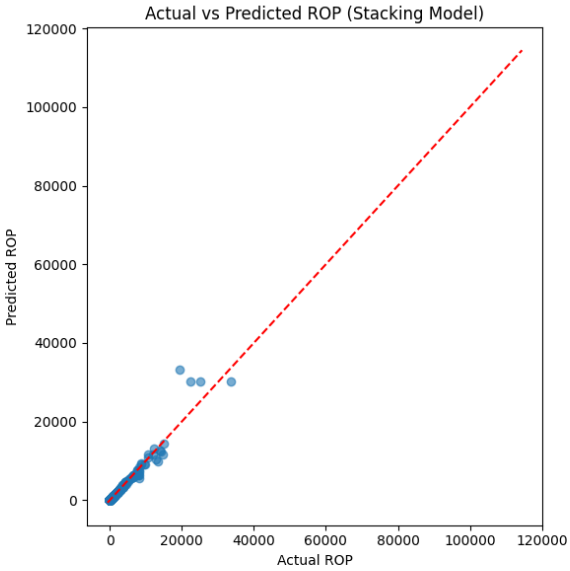


Figure 3.10: Actual vs Predicted Reorder Point (ROP).

These results show that the stacking ensemble produced the best overall performance, having lowest RMSE and MAE with a very high R^2 value.

3.6.5 Visualization of Model Predictions

The stacking model performance could be interpreted visually through a scatter plot comparison between actual ROP values and predicted ROP values. Points that closely follow the diagonal line indicate successful model operation (Shao et al., 2017). The visual evidence showed that most predicted values closely followed actual measurement results which validated the statistical assessment.

3.7 Model Deployment

The final phase of the research involved deploying the predictive and optimization models into a user-accessible system. A web-based interface was developed using Streamlit (Richards, 2021) to allow pharmacists and procurement officers to interact with the system without requiring programming expertise.

The deployment served three main purposes: transforming predictive analytics into a practical decision-support tool, enhancing user understanding through built-in visualizations, and enabling evidence-based stock and procurement planning.

While this section introduces the deployment strategy, further details on the system's architecture and logic flow are provided in Chapter 4 under system design.

3.7.1 Model Maintenance and System Integration

To ensure long-term reliability and usability, the predictive system should incorporate both model maintenance and seamless integration with hospital workflows. A retraining strategy should be established to monitor model performance over time and trigger retraining quarterly or when accuracy metrics such as R^2 fall below a threshold (e.g., 0.75). Tools like DVC or MLflow can support model versioning, data lineage tracking, and reproducibility.

In parallel, the model can be integrated into the hospital information system (HIS) via RESTful APIs or lightweight middleware services. This real-time linkage will allow automatic synchronization of inventory data and timely forecasting, thereby reducing manual entry errors and improving responsiveness in reorder decisions. Together, these strategies ensure the system remains accurate, scalable, and operationally aligned with real-world healthcare needs.

Chapter 4

System Design and Architecture

4.1 Introduction

This chapter describes the architectural design of the pharmaceutical inventory optimization and reorder point (ROP) prediction system. The system design presents an extensive outline of its organizational structure together with the relationships between its different functional units. The architectural design determines the data process from raw inputs to model predictions and visualization elements to create a smooth experience for all hospital staff who work with pharmaceutical inventory management.

4.2 System Requirements

The system accepts different pharmaceutical data types needed by hospital pharmacists and procurement officers for average demand information and demand variability along with stock turnover rate inputs. The system feeds its data to a stacking ensemble machine learning model that uses historical data for effective reorder point prediction. The system further incorporates an optimization component that considers ordering, shortage costs, overstocking costs, and holding costs to determine the most efficient stock levels. Streamlit serves as the framework for the user interface, while Plotly powers the interactive graphics. Additional libraries, such as `scikit-learn` and `joblib`, facilitate model loading, preprocessing, and prediction tasks.

4.2.1 Model Execution Requirements

The system begins with necessary preprocessing before moving on to prediction execution. The model requires numerical inputs like mean demand and standard deviation to undergo scaling that matches the training parameters of the model. The model requires categorical features to undergo encoding because this helps it understand their meaning correctly. The stacking model performs its prediction task after which it combines the results with additional input data. The system combines the prediction process with cost-based optimization to deliver clinical recommendations that also meet financial requirements.

4.3 Overview of System Architecture

The system divides its functionality into three distinct layers called data handling and predictive modeling and output visualization. The system obtains data input from CSV files through data handling when dealing with inventory analytics but uses a user-friendly form interface to process single-product predictions. The predictive modeling phase depends on a stacking ensemble regressor that analyzes user inputs together with historical demand patterns. An optimization process follows prediction generation to assess holding and ordering expenses for optimizing safety stock quantities and order size. The visual components of the display utilize the Streamlit frontend to show the final outputs.

4.4 Frontend Development

4.4.1 User Interface Design

Streamlit serves as the foundation for the user interface through which users interact with a basic form for data entry. A single-product prediction requires users to manually supply essential information including mean demand statistics and demand variability as well as stock turnover data. After input data submission the system preprocesses model data before delivering the reorder point output to the user interface. Hospital staff members can access predictions through a web-based frontend without needing any specialized software to operate on their local computers. The system presents its second page which serves as a platform for viewing broader inventory dashboards. The preprocessed CSV file automatically produces multiple comparative metrics for different products.

4.4.2 Visualization from the Optimization Model

The system generates different visual aids after predictions from the model and optimization procedures are finished. The bar charts display information about predicted reorder points together with optimal inventory levels. Scatter plots display the connections between stock turnover rates and current inventory amounts as well as reorder point levels. KPIs shown on the dashboard display cost savings together with turnover rates and projected stockout risks. Clinical and procurement teams use visual representations to track inventory health which enables them to take appropriate corrective actions.

4.5 Backend Development

4.5.1 Model Integration

The stacking ensemble model gets saved through `joblib` for use in the live system at execution time. The system applies standardization procedures to numeric fields and uses the same encoding process for categories after the user enters data. This encoding process matches the approach used during model development. The system runs the model to determine the suitable reorder point. The application provides a seamless workflow that enables hospital staff to easily perform preprocessing and make final predictions so they need little technical supervision.

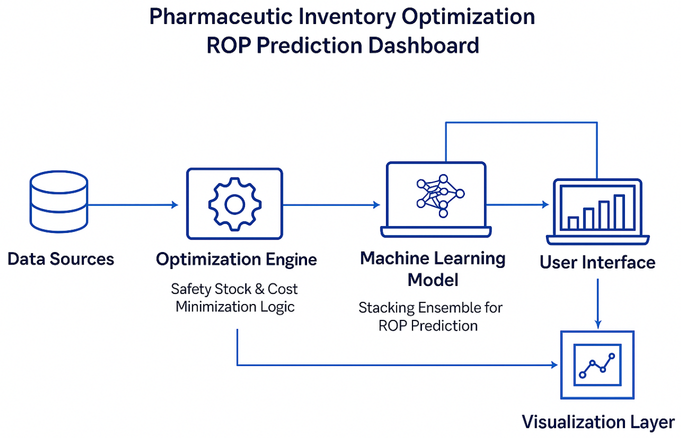


Figure 4.1: System Architecture

4.5.2 Optimization Computation

Beyond predicting reorder points, the backend runs additional computations to determine safety stock and order quantities. It applies a cost function that weighs holding costs, ordering expenses, and potential wastage or shortages, with the goal of recommending an optimal level of inventory. Once these figures are calculated, they are sent to the interface for real-time presentation, enabling pharmacy managers to balance cost efficiency against the risk of running low on critical supplies.

The system design leverages modular architecture and a well-defined data flow. The frontend component enables rapid product data entry while presenting accessible charts and KPIs for analyzing the final results to hospital staff. The single pipeline within the backend consists of preprocessing steps alongside prediction through the stacking ensemble model and cost-based optimization calculations. A combined approach of effective user interface design and advanced prediction methods within the system allows healthcare personnel to access valuable cost-saving insights for pharmaceutical inventory decisions.

Chapter 5

System Implementation and Testing

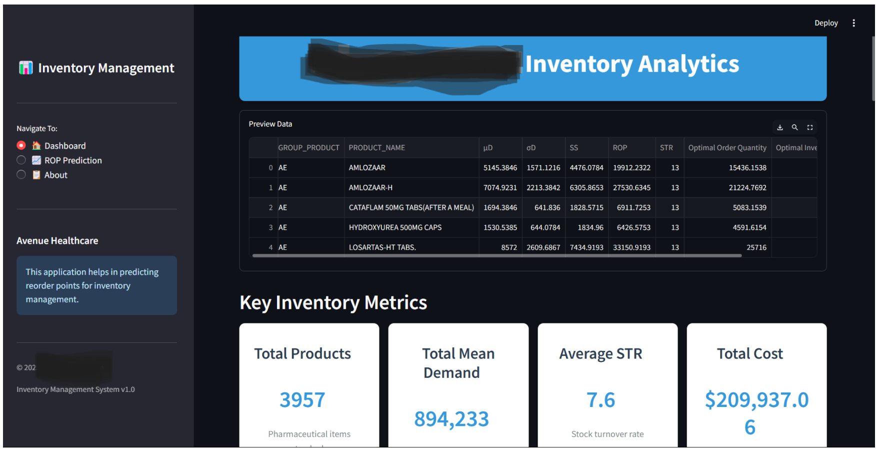
5.1 Introduction

This chapter outlines the implementation of the predictive analytics system and discusses key considerations for future validation, usability, and fairness. Although the system was not tested with hospital staff during this research, design decisions prioritized usability and clinical relevance to enable future adoption in healthcare settings.

5.2 User Interface Design

The hospital staff can access the interface through Streamlit which provides a basic browser platform. The application reveals three main pages which users can access through the sidebar after starting the program. The introductory page of the application provides essential information about ABC-VEN classes and system functions to viewers. The second page—*ROP Prediction*—allows users to input specific product information (e.g., mean demand, standard deviation of demand, turnover rate, and safety stock). The machine learning routine executes after users submit the form to generate the recommended reorder point calculation. The system displays both the calculated ROP value and produces optional charts and downloadable CSV files for offline use during this step. The third page named *Analytics Dashboard* retrieves existing data from the server file to present multiple product analytics including cost distribution and reorder point statistics and additional metrics.

Single-product forecasting is possible through the ROP Prediction page. Users complete straightforward fields called “Mean Demand” and “Standard Deviation” in the form. The



Inventory Management

Navigate To:
● Dashboard
○ ROP Prediction
○ About

Avenue Healthcare

This application helps in predicting reorder points for inventory management.

© 202[redacted]

Inventory Management System v1.0

[redacted] Inventory Analytics

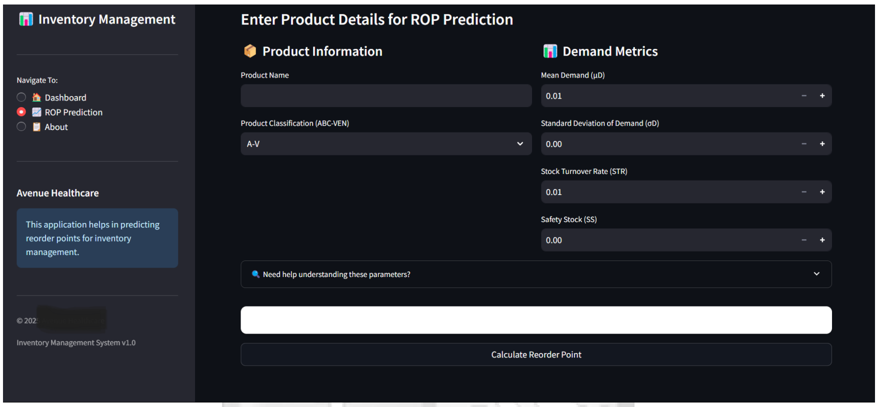
Preview Data

	GROUP_PRODUCT	PRODUCT_NAME	µD	σD	SS	ROP	STR	Optimal Order Quantity	Optimal Inve
0	AE	AMLOZAAR	5145.3846	1571.1216	4476.0784	19912.2322	13	15436.1538	
1	AE	AMLOZAAR-H	7074.9231	2213.3842	6305.8653	27530.6345	13	21224.7692	
2	AE	CATAFLAM 50MG TABS(AFTER A MEAL)	1694.3846	641.836	1828.5715	6911.7253	13	5083.1539	
3	AE	HYDROXYUREA 500MG CAPS	1530.5385	644.0784	1834.96	6426.5753	13	4591.6154	
4	AE	LOSARTAS-HT TABS.	8572	2609.6867	7434.9193	33150.9193	13	25716	

Key Inventory Metrics

Total Products	Total Mean Demand	Average STR	Total Cost
3957	894,233	7.6	$209,937.06
Pharmaceutical items		Stock turnover rate	

Figure 5.1: Sample overview of the web application



Inventory Management

Navigate To:
○ Dashboard
● ROP Prediction
○ About

Avenue Healthcare

This application helps in predicting reorder points for inventory management.

© 202[redacted]

Inventory Management System v1.0

Enter Product Details for ROP Prediction

Product Information

Product Name:

Product Classification (ABC-VEN): A-V

Demand Metrics

Mean Demand (µD): 0.01

Standard Deviation of Demand (σD): 0.00

Stock Turnover Rate (STR): 0.01

Safety Stock (SS): 0.00

Figure 5.2: ROP Prediction page

step allows users to quickly obtain an ROP through simple numerical value input. The system produces error messages and warning alerts whenever users enter unusual or incorrect inputs into the fields. This prompts users to check their data entries.

5.3 Model Integration and Execution

The system uses a stacking ensemble model for its predictive operations. When the application launches the model becomes available to perform reorder point calculations swiftly. The system validates ROP Prediction form inputs by checking mean demand and standard deviation and turnover rate values meet basic criteria and follow training preprocessing methods. The system suggests a reorder point recommendation that it can improve through safety stock estimation techniques and cost-effective ordering algorithms to assist users in their decision-making process.

The system produces brief analytics after the initial prediction by showing charts which display the user-entered safety stock level alongside the newly computed reorder point. A bar chart shows both safety stock and predicted reorder point values with clear visual distinction between these numbers. The user receives guidance through additional tables and bullet points to understand risk management approaches such as order placement timing and assessment of abnormal reorder point levels compared to mean demand.

5.4 Analytics Dashboard

Users access single-product forecasting through the *ROP Prediction* page but the system also provides the *Analytics Dashboard* which features three distinct graphical views like a bar chart of top ten product costs and scatter and radar charts illustrating inventory data.

The dashboard presents information about many products simultaneously to display inventory status along with restocking patterns and associated pricing effects across all products. Hospital administrators can use single-product on-demand forecasting through the *ROP Prediction* page combined with the dashboard for analyzing multiple products in aggregate.

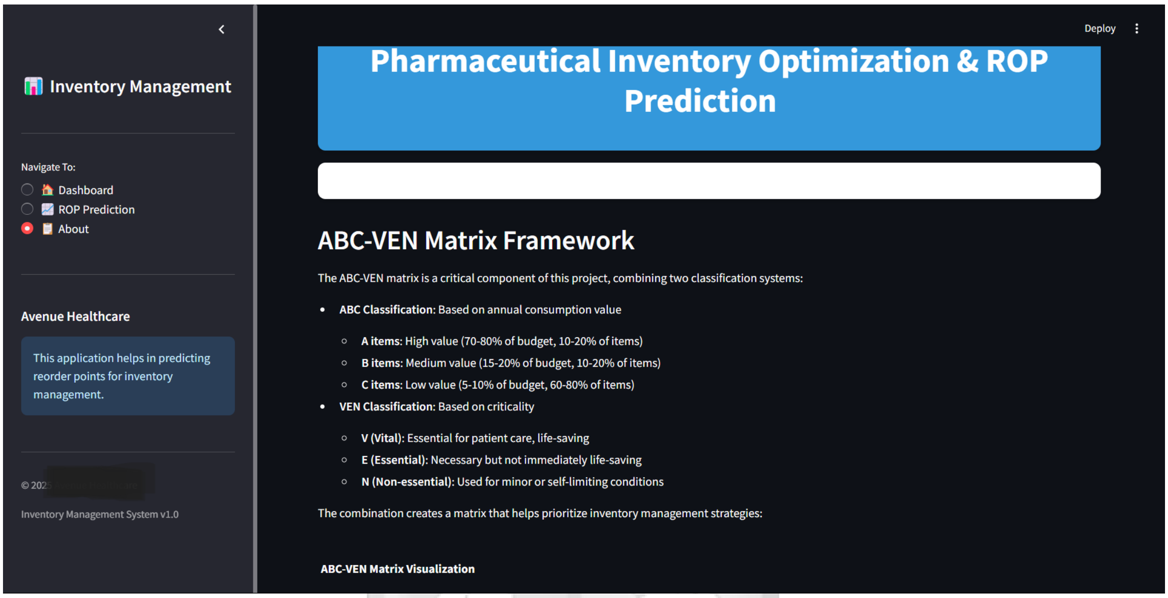
This application helps in predicting reorder points for inventory management.

Figure 5.3: Dashboard for the Analytics

The pharmaceutical inventory optimization and ROP prediction system received an implementation focused on healthcare staff usability to prevent limited technical expertise from blocking its advantages. The system design features single-product predictions in a separate module which serves daily restocking requirements and also supports strategic planning activities. The system proved to generate accurate reorder point forecasts and interactive analysis tools which enhance inventory management efficiency according to testing results. Performance metrics resulting from the study appear in the following chapter, together with an assessment of system effectiveness and a list of necessary future work.

5.5 ROP Results

Once the data is validated and processed, the machine learning model makes predictions for the reorder points (ROP) based on the input features. The predicted reorder points are displayed in a table showing the product name, product group, and the predicted reorder point (ROP). This table allows the user to quickly review the results and make decisions regarding inventory management.



This application helps in predicting reorder points for inventory management.

© 202[redacted]

Inventory Management System v1.0

Pharmaceutical Inventory Optimization & ROP Prediction

ABC-VEN Matrix Framework

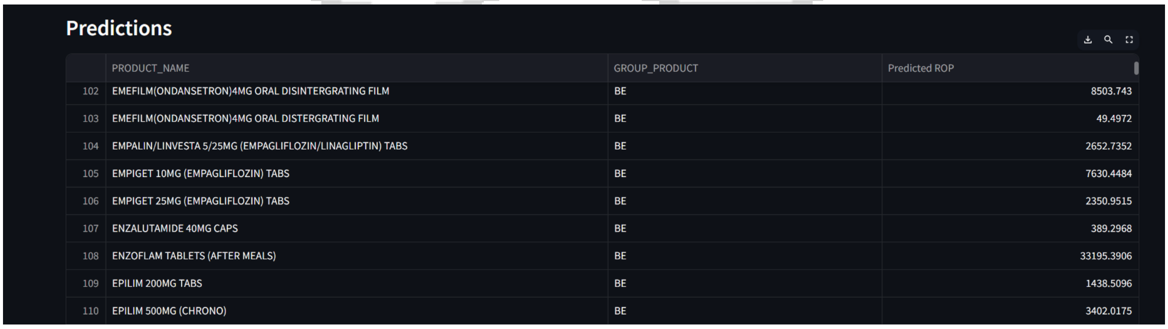
The ABC-VEN matrix is a critical component of this project, combining two classification systems:

- **ABC Classification:** Based on annual consumption value
  - **A items:** High value (70-80% of budget, 10-20% of items)
  - **B items:** Medium value (15-20% of budget, 10-20% of items)
  - **C items:** Low value (5-10% of budget, 60-80% of items)
- **VEN Classification:** Based on criticality
  - **V (Vital):** Essential for patient care, life-saving
  - **E (Essential):** Necessary but not immediately life-saving
  - **N (Non-essential):** Used for minor or self-limiting conditions

The combination creates a matrix that helps prioritize inventory management strategies:

ABC-VEN Matrix Visualization

Figure 5.4: The About page



Predictions

	PRODUCT_NAME	GROUP_PRODUCT	Predicted ROP
102	EMEFILM(ONDANSETRON)4MG ORAL DISINTERGRATING FILM	BE	8503.743
103	EMEFILM(ONDANSETRON)4MG ORAL DISTERGRATING FILM	BE	49.4972
104	EMPALIN/LINVESTA 5/25MG (EMPAGLIFLOZIN/LINAGLIPTIN) TABS	BE	2652.7352
105	EMPIGET 10MG (EMPAGLIFLOZIN) TABS	BE	7630.4484
106	EMPIGET 25MG (EMPAGLIFLOZIN) TABS	BE	2350.9515
107	ENZALUTAMIDE 40MG CAPS	BE	389.2968
108	ENZOFLAM TABLETS (AFTER MEALS)	BE	33195.3906
109	EPILIM 200MG TABS	BE	1438.5096
110	EPILIM 500MG (CHRONO)	BE	3402.0175

Figure 5.5: ROP predictions

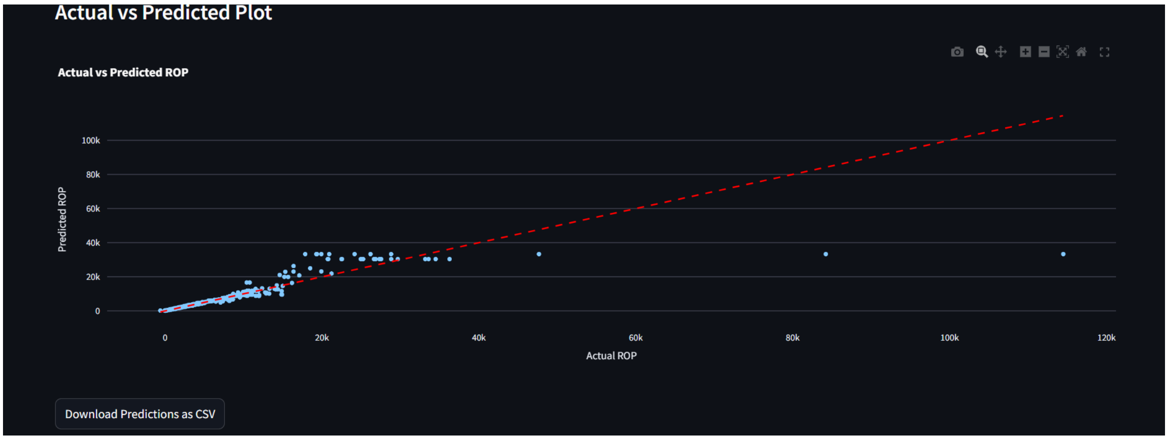


Figure 5.6: Model Generalization

Additionally, a scatter plot is generated comparing the actual ROP (if available in the dataset) against the predicted ROP. The plot includes a red diagonal line representing perfect predictions (where the actual ROP equals the predicted ROP), helping the user assess the accuracy of the predictions. Users can also download the predicted ROP values in a CSV format for further analysis or integration with inventory management systems.

Chapter 6

Results and Discussions

6.1 Introduction

The research findings about the pharmaceutical inventory optimization together with reorder point (ROP) prediction system form the basis of this chapter. The discussion examines ROP prediction effectiveness together with ML model performance alongside the visual outputs from the optimization model. The results interpretation of this chapter reveals essential insights about the system to help hospital procurement officers and pharmacists enhance their inventory management practices.

6.2 Analysis of ROP Predictions

The pharmaceutical inventory optimization system performs forecasting of reorder points (ROP) through predictive modeling and historical sales data analysis for each product. These predictions require high accuracy since they shape both inventory management choices and procurement approaches.

The Chapter 5 Actual vs Predicted ROP scatter plot demonstrates the proximity between real reorder points and their forecasted values. The prediction values should appear directly beside the actual ROP values which are shown through the red diagonal line. The model shows strong correlation yet it produces some errors when predicting reorder points for products with lower demand levels. The prediction accuracy metrics R^2 demonstrate promising results although some slight inaccuracies occur in specific product categories.

6.3 Machine Learning Model Performance

The developed stacking ensemble model from this study combines multiple learning algorithms to generate an accurate prediction of reorder points (ROPs). The stacking ensemble model linked Random Forest and Gradient Boosting ensemble methods with a linear meta-model to achieve effective bias-variance balance which ensures generalized predictions for new data points.

The model developed was evaluated against three measures of performance that included the R^2 coefficient of determination, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The model explains 84% of the variability in reorder points through its input features according to the R^2 value. The strong prediction correlation between the input variables and output is retained because the value of R^2 is maintained within the intended threshold, avoiding overfitting.

With an RMSE value of 103 the best model demonstrated that major discrepancies between predicted and actual ROPs remained small. The average deviation of 20 units in MAE shows that predicted reorder points were similar to actual values. The system requires such accurate forecasting because healthcare inventory systems need to avoid under- and over-stocking situations that create substantial service and cost challenges.

The accuracy and reliability of stacking ensemble models have been confirmed through these results for inventory management decision support. The precise ROP estimates generated by this model provide timely information to procurement officers and pharmacists who use it to improve their stock control processes and enhance essential medicine availability.

6.4 Insights from the Optimization Model Results

The optimization model functions alongside machine learning predictions to deliver valuable inventory management insights. The optimization engine performs calculations of optimal inventory levels and order quantities after receiving the reorder points predictions along with safety stock and cost minimization logic.

The optimization model visual output reveals how the best order quantities relate to best inventory levels across different products. Figure 6.1 presents the scatter plot which displays this relationship to help procurement officers determine necessary replenishment attention for different products. The products that fall in higher ranges of optimal order quantity and optimal inventory level require immediate attention to adjust inventory.

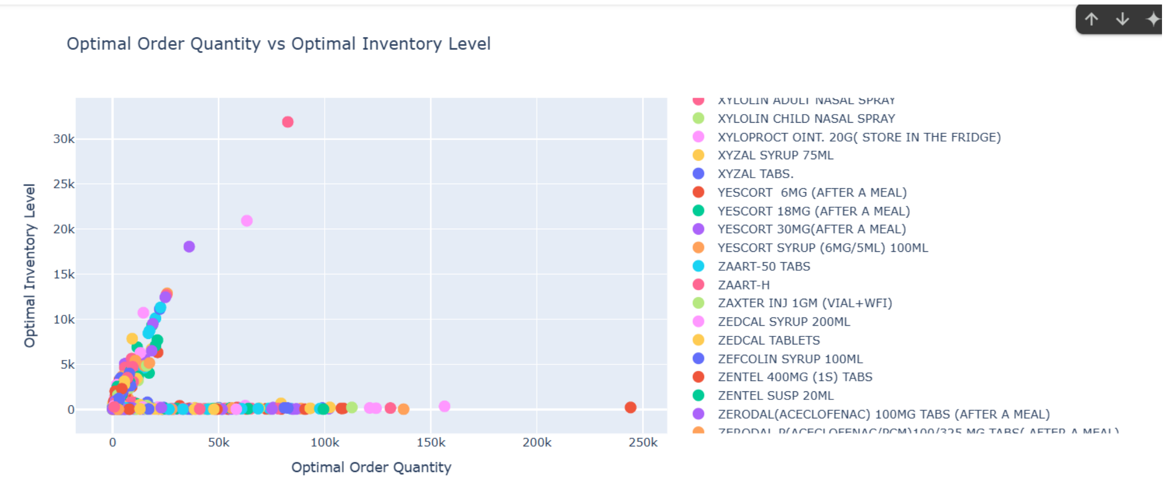


Figure 6.1: Optimal Order Quantity vs Optimal Inventory Level

By displaying the optimal order quantity against optimal inventory level on a scatter plot procurement teams can easily determine which products need restocking and in what quantities. The team can rapidly determine products requiring immediate attention by observing the point distribution and color-coded product names which show large optimal order quantity and current inventory level gaps.

The optimization model effectively manages predicted reorder points together with safety stock and order quantities making it a significant benefit of the system. The system provides decision transparency through display of scatter plots and bar charts and KPIs which demonstrate how different inventory levels and order quantities affect costs.

The bar chart named **Top 10 Products - Reorder Point (ROP)** provides predicted reorder point data for the ten most crucial products and indicates their importance based on their ROP values.

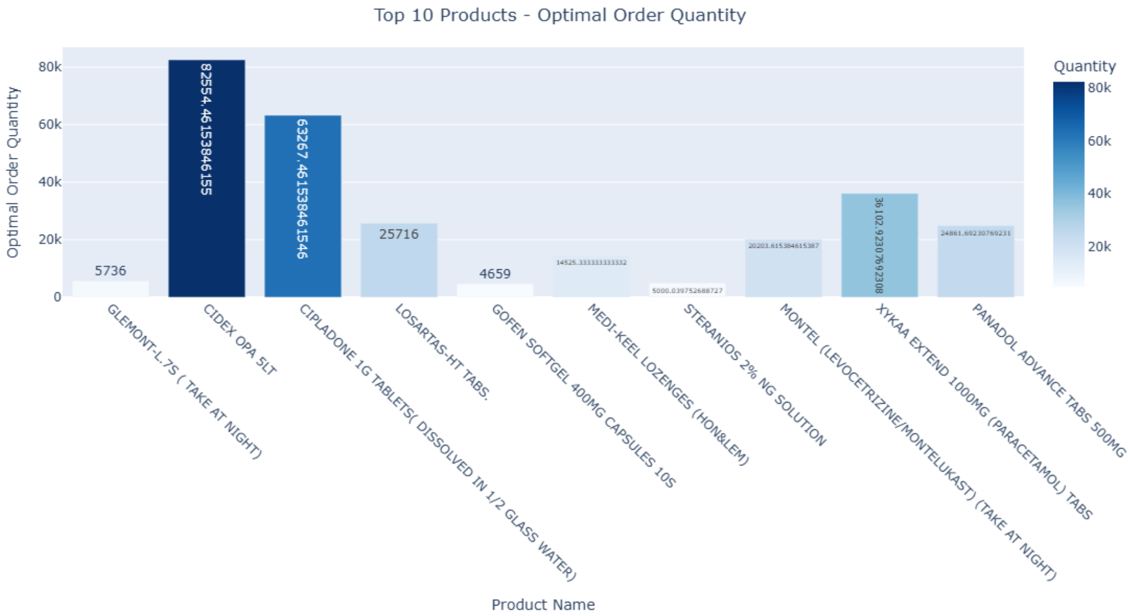


Figure 6.2: Top 10 Products - Reorder Point (ROP)

The bar chart shows how much total cost each product category incurs through analysis of its inventory levels and order quantity. Better cost management strategies should be developed for products that exhibit high costs according to this visual representation.

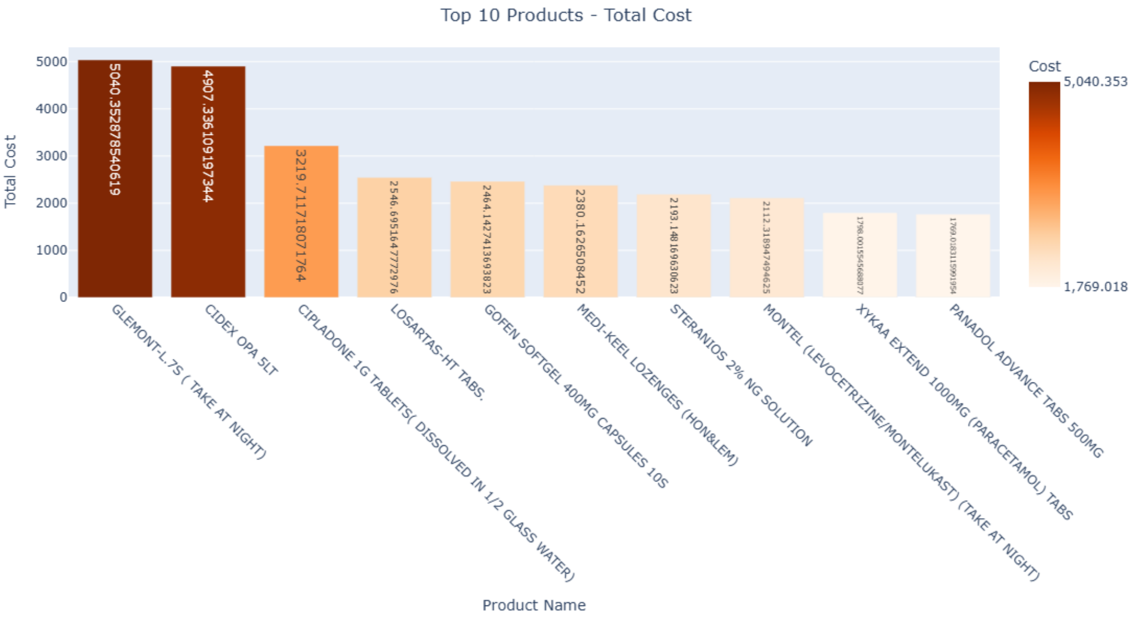


Figure 6.3: Top 10 Products - Total Cost

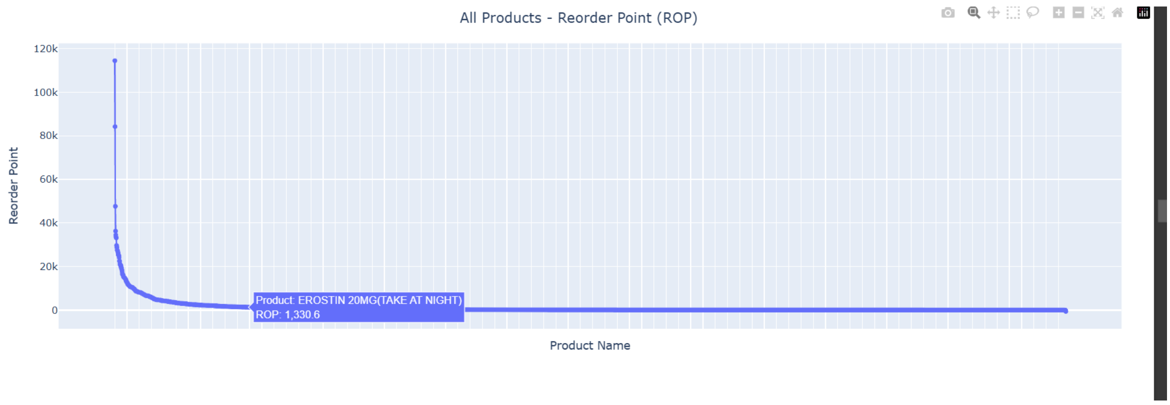


Figure 6.4: All Products - Reorder Point (ROP)

6.5 Visualization of All Products - Reorder Point (ROP)

Figure 6.4 displays the **line plot** containing all products' reorder points. The plot shows how different products have widely varying reorder points where several products need very high reorder points yet others require much lower levels. The plot provides procurement officers with important data to determine restocking frequency by showing which products reach their reorder points thus enabling them to focus on popular items in need of immediate attention.

Among all products in the plot "EROSTIN 20MG (TAKE AT NIGHT)" requires the highest reorder point. Frequent restocking of the product becomes necessary because its demand levels remain high or its supply lead times extend beyond other products. The identification of these patterns helps maximize inventory management by maintaining essential products ready for use.

The Figure 6.5 presents **Reorder Points (ROP)** for **Top 10 Products** that show large stock requirement differences. The ROP of **Glemont-L.7S (Take at Night)** stands at the peak among all products thus demonstrating urgent restocking needs due to its high demand. The reorder point for **CIDEX OPA 5LT** stands out as high because of its extensive usage needs and **Cipladone 1G Tablets** and **Steranos 2% NG Solution** exhibit lower reorder points indicating their lower importance. The distribution pattern shows two distinct peaks because some items need large stock levels because of heavy demand but other products function well with reduced inventory requirements. The understanding obtained from this analysis

Figure 6.5: Top 10 Products - Reorder Point (ROP)

enables procurement officers to determine product priorities according to their usage levels and establish inventory quantities that prevent both stockout situations and overstocking conditions.

These visual outputs present procurement officers with clear directions about which restocking priorities exist along with strategies to minimize inventory expenses.

The optimization model achieves its strength by maintaining equilibrium between future reordering points and safety stock levels and order quantity decisions. Decision-making transparency occurs through the bar charts and KPIs that display cost information about different order quantities and inventory levels.

The system's dual capability to reduce costs and maintain prompt restocking operations enables better inventory efficiency. The model determines optimal order quantities which the system uses to compare them against current inventory levels to prevent both stockouts and overstocking issues that waste resources and optimize procurement budgets.

6.6 Practical Implications for Healthcare Procurement

The combination of the ML model and the optimization engine in this system gives a revamped inventory management facility to hospital procurement officers. Hospital inventory replenishment by timing is made optimal using the system which utilizes precise reorder point forecasts. Hospitals can help prevent costs arising from overstocking and out-of-inventory conditions by embracing this system.

Besides, the system supports data-driven procurement, and real-time optimization results dynamically vary inventory levels in accordance with demands. Such benefits translate to a more efficient and cost-saving inventory management system that can maximize the availability of necessary pharmaceuticals within hospitals, eventually resulting in better patient care.

Chapter 7

Conclusion and recommendations

7.1 Conclusion

In this research work, we managed to design and deploy a pharmaceutical reorder point (ROP) forecasting and inventory optimization system. Using machine learning algorithms, we utilized combined historical sales data, lead times, and inventory levels to project optimal reorder points for medicines. We deployed the system through Python programming language and the Streamlit framework as the frontend, presenting an accessible interface for hospital procurement officers and pharmacists.

The model was extensively pre-processed, including feature scaling and encoding categorical data, to set it up for prediction. With the model trained on a full dataset, it performed well in predicting the reorder points, with promising accuracy in terms of R^2, RMSE, and MAE measures. These metrics of performance indicate that the model is reliable for real-world application, providing precise reorder point predictions that can be used to inform inventory management choices.

The inclusion of an optimization engine in the machine learning model enabled the system to calculate not only the reorder points but also optimal inventory levels and order quantities, taking into account safety stock and cost minimization rules. The graphical outputs, including scatter plots, bar charts, and line plots, make it easy for procurement officers to assess inventory requirements and make sound decisions.

With the optimization of reorder points and stock levels, the system allows hospitals to manage pharmaceutical stocks more efficiently, minimizing stockouts and overstocking and saving money. The application in real-world scenarios demonstrates the ability of AI and machine learning to improve supply chain management in healthcare settings.

The success of this project illustrates the transformative potential of AI in optimizing pharmaceutical inventory management. By providing accurate predictions and actionable insights, the system enables more effective procurement, which means improved availability of life-saving medicines in healthcare facilities and better patient outcomes.

7.2 Limitations and Areas for Improvement

While the system has been promising, there are some limitations and points of improvement. The quality of the input data plays a big role in determining the performance of the system. Incorrect or incomplete data can lead to erroneous predictions, and later releases of the system can include more advanced data validation checks and data cleaning techniques.

Although the system works well with the dataset provided, further testing is necessary to ensure that the model generalizes effectively to other types of pharmaceutical inventory, particularly in different healthcare settings. Additionally, incorporating real-time sales and inventory data could further improve the accuracy of reorder point predictions and allow for dynamic inventory adjustments based on current demand.

7.3 Future Work

While the system developed in this study has encouraging results, there are several ways that it can be enhanced in the future. One of the most significant is the inclusion of more diverse and comprehensive datasets and multi-institutional datasets. For example, including storage data for types of drugs such as cost per square meter of storage area and maintaining up-to-date storage conditions up to date would significantly enhance the formulation of the optimization model. This would allow the system to consider more precise factors in making decisions, giving better optimization of stock handling in space and storage conditions.

In addition, the system would be strengthened by incorporating up-to-date sales data and inventory levels, which would make the reorder point estimates more real-time and responsive to current conditions. Real-time data would improve the flexibility of the system in reacting to real-time demand changes, making it a more powerful system for continuous inventory control.

7.4 Recommendations

To realize maximum possible utility and efficiency of this pharmaceutical inventory system, certain suggestions are made. Firstly, collaborating with technology companies and research institutes can help synchronize the system into the existing medical infrastructure. Such collaboration can introduce the technical backup and resources necessary to scale the system to be utilized in healthcare institutions on a wider scale.

In addition, medical staff needs to be trained on how to use the system effectively. Proper training ensures that procurement officers, pharmacists, and other medical officers understand the functionality of the system and can use it to improve the management of inventories.

Engaging with policymakers could also help in the creation of supportive guidelines and regulations for the widespread adoption of such AI-driven tools in healthcare. These guidelines can help standardize the application of predictive models and ensure that they align with best practices in medical supply chain management.

Lastly, establishing monitoring and feedback systems is necessary in order to improve the system continuously. This would give periodic updates and revisions based on users' feedback and also real-world problems, so that the system can be efficient and relevant in the long term.

References

Abu Zwaida, T., Pham, C., and Beauregard, Y. (2021a). Optimization of inventory management to prevent drug shortages in the hospital supply chain. *Applied Sciences*, 11(6):2726.

Abu Zwaida, T., Pham, C., and Beauregard, Y. (2021b). Optimization of inventory management to prevent drug shortages in the hospital supply chain. *Applied Sciences*, 11(6).

Adeghe, E. P., Okolo, C. A., and Ojeyinka, O. T. (2024). A review of the use of machine learning in predictive analytics for patient health outcomes in pharmacy practice. *OARJ of Life Sciences*, 7(1):052–058.

Ahrens, A., Ersoy, E., Iakovlev, V., and Li, H. (2022). An introduction to stacking regression. *Credible Asset Allocation, Optimal Transport Methods, and Related Topics*, 429:7.

Al-Nsour, L. H. (2024). Optimizing hospital pharmacy inventory: A robust model predictive control approach. Master's thesis, State University of New York at Binghamton.

Ali, A. K. (2011). Inventory management in pharmacy practice: a review of literature. *Archives of pharmacy practice*, 2(4-2011):151–156.

Angula, T. N. and Dongo, A. (2024). Assessing the impact of artificial intelligence and machine learning on forecasting medication demand and supply in public pharmaceutical systems: a systematic review. *GSC Biological and Pharmaceutical Sciences*, 26(2):140–150.

da Costa, L. S. (2023). *Applying Artificial Intelligence to Improve the Effectiveness of the Demand Forecasting in a Specialty Pharmacy*. PhD thesis, University of Cincinnati.

Deressa, M. B., Beressa, T. B., and Jemal, A. (2022). Analysis of pharmaceuticals inventory management using abc-ven matrix analysis in selected health facilities of west shewa zone, oromia regional state, ethiopia. *Integrated Pharmacy Research and Practice*, pages 47–59.

Di Bucchianico, A. (2008). Coefficient of determination (r 2). *Encyclopedia of statistics in quality and reliability*.

Ebojoh, E., Jewo, A., and Ogbebor, O. (2021). Calculating eoq and rop for some equipment. *NIPES-Journal of Science and Technology Research*, 3(1).

Emmert-Streib, F. and Dehmer, M. (2019). Evaluation of regression models: Model assessment, model selection and generalization error. *Machine learning and knowledge extraction*, 1(1):521–551.

Fernandez, R. M. (2020). Sdg3 good health and well-being: integration and connection with other sdgs. *Good Health and Well-Being*, pages 629–636.

Galli, L., Levato, T., Schoen, F., and Tigli, L. (2021). Prescriptive analytics for inventory management in health care. *Journal of the Operational Research Society*, 72(10):2211–2224.

Hodson, T. O. (2022). Root mean square error (rmse) or mean absolute error (mae): When to use them or not. *Geoscientific Model Development Discussions*, 2022:1–10.

Karamshetty, V., De Vries, H., Van Wassenhove, L. N., Dewilde, S., Minnaard, W., Ongarora, D., Abuga, K., and Yadav, P. (2022). Inventory management practices in private healthcare facilities in nairobi county. *Production and Operations Management*, 31(2):828–846.

Kirimi, M. J. (2023). *Inventory Management Practices and Public Health Institutions Performance in Nairobi City County, Kenya*. PhD thesis, Kenyatta University.

Kumar, V. R., Waghmare, P., Bukya, S., Depuru, B. K., and Kaliamoorthy, I. (2023). Forecasting drug demand for optimal medical inventory management: A data-driven approach with advanced machine learning techniques. *International Journal of Innovative Science and Research Technology*, 8(9):221–229.

Magarsa Bayissa Deressa, T. B. B. and Jemal, A. (2022). Analysis of pharmaceuticals inventory management using abc-ven matrix analysis in selected health facilities of west shewa zone, oromia regional state, ethiopia. *Integrated Pharmacy Research and Practice*, 11:47–59. PMID: 35242694.

Mawengkang, H. et al. (2020). An optimization model for hospitals inventory management in pharmaceutical supply chain. *Systematic Reviews in Pharmacy*, 11(3).

Mohammed, S. A. and Workneh, B. D. (2020). Critical analysis of pharmaceuticals inventory management using the abc-ven matrix in dessie referral hospital, ethiopia. *Integrated Pharmacy Research and Practice*, pages 113–125.

Mousa, B. A. and Al-Khateeb, B. (2023). Predicting medicine demand using deep learning techniques: A review. *Journal of intelligent systems*, 32(1):20220297.

MSH (2012). *MDS-3: Managing Access to Medicines and Health Technologies*. Management Sciences for Health.

Musimbi, P. M. (2022). *A Predictive analytics model for pharmaceutical inventory management.* PhD thesis, Strathmore University.

Pall, R., Gauthier, Y., Auer, S., and Mowaswes, W. (2023). Predicting drug shortages using pharmacy data and machine learning. *Health Care Management Science*, 26(3):395–411.

Quick, J. D. (1982). Applying management science in developing countries: Abc analysis to plan public drug procurement. *Socio-Economic Planning Sciences*, 16(1):39–50.

Richards, T. (2021). *Getting Started with Streamlit for Data Science: Create and deploy Streamlit web applications from scratch in Python*. Packt Publishing Ltd.

Schröer, C., Kruse, F., and Gómez, J. M. (2021). A systematic literature review on applying crisp-dm process model. *Procedia Computer Science*, 181:526–534.

Shao, L., Mahajan, A., Schreck, T., and Lehmann, D. J. (2017). Interactive regression lens for exploring scatter plots. In *Computer graphics forum*, volume 36, pages 157–166. Wiley Online Library.

Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., and Müller, K.-R. (2021). Towards crisp-ml (q): a machine learning process model with quality assurance methodology. *Machine learning and knowledge extraction*, 3(2):392–413.

Toroitich, A. M., Dunford, L., Armitage, R., and Tanna, S. (2022). Patients access to medicines–a critical review of the healthcare system in kenya. *Risk Management and Healthcare Policy*, pages 361–374.

Willmott, C. J. and Matsuura, K. (2005). Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82.

Appendix A

Similarity Index

15% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

182 Not Cited or Quoted 13%
Matches with neither in-text citation nor quotation marks

15 Missing Quotations 1%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

9%	Internet sources
7%	Publications
12%	Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

182 Not Cited or Quoted 13%
Matches with neither in-text citation nor quotation marks

15 Missing Quotations 1%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

9%	Internet sources
7%	Publications
12%	Submitted works (Student Papers)

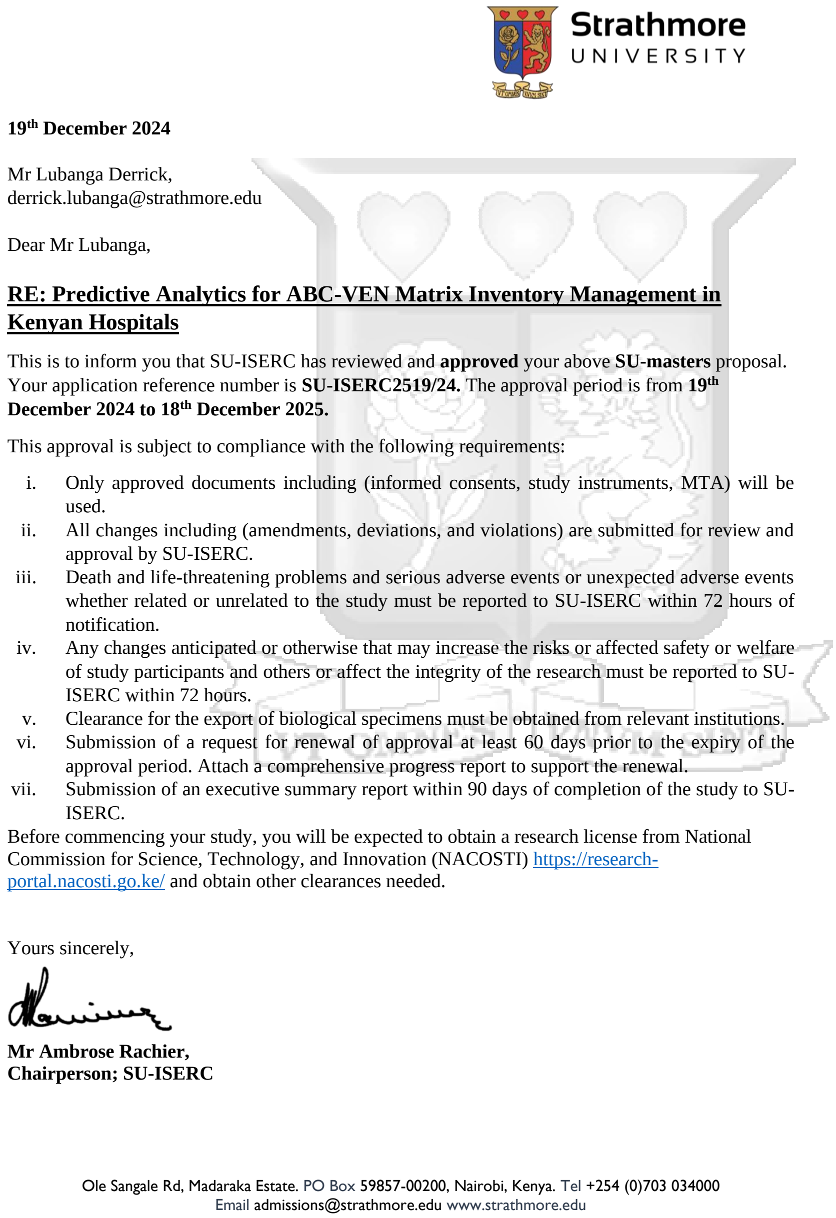
Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Submitted works	**Columbia College of Missouri on 2024-03-06**	<1%
2	Internet	**su-plus.strathmore.edu**	<1%
3	Internet	**www.coursehero.com**	<1%
4	Submitted works	**Centre for Distance and Online Education Galgotias University on 2024-12-23**	<1%
5	Internet	**www.mdpi.com**	<1%
6	Internet	**etd.aau.edu.et**	<1%
7	Internet	**ijisrt.com**	<1%
8	Submitted works	**Nottingham Trent University on 2025-03-13**	<1%
9	Submitted works	**University of Liverpool on 2013-11-07**	<1%
10	Internet	**pharmacyeducation.fip.org**	<1%

Appendix B

Ethical approval



Strathmore
UNIVERSITY

19th December 2024

Mr Lubanga Derrick,
derrick.lubanga@strathmore.edu

Dear Mr Lubanga,

<u>RE: Predictive Analytics for ABC-VEN Matrix Inventory Management in Kenyan Hospitals</u>

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2519/24.** The approval period is from **19th December 2024 to 18th December 2025.**

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-ISERC.
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-ISERC within 72 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-ISERC within 72 hours.
v. Clearance for the 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 the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before 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,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu

Appendix C

Python Code for Optimization and Prediction

Appendix C lists all the Python scripts used to aggregate demand, clean data, optimize inventory, and machine learning-based reorder point prediction. The code is put into logical divisions to make it understandable.

C.1 A.1 Data Cleaning and Aggregation

```python
import pandas as pd
import matplotlib.pyplot as plt
import plotly.graph_objects as go
import numpy as np
import seaborn as sns
from statsmodels.tsa.arima.model import ARIMA
import plotly.express as px
import plotly.subplots as sp
import warnings
from statsmodels.tsa.stattools import adfuller
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
warnings.filterwarnings("ignore")

file_path = "/content/drive/MyDrive/Colab Notebooks/ABC-VEN/Avenue Healthcare - Sales Dump - Year_2023.xlsb"
sheet_names = pd.ExcelFile(file_path).sheet_names

```

```
def fix_date_format(date_column):
    try:
        fixed_dates = pd.to_datetime(date_column, errors='coerce', format='%m/%d/%Y')
    except:
        fixed_dates = None
    if fixed_dates.isnull().all():
        try:
            fixed_dates = pd.to_datetime(pd.to_numeric(date_column, errors='coerce'), origin='31-12-1899', unit='D')
        except:
            fixed_dates = None
    return fixed_dates

sales_df = pd.concat([
    pd.read_excel(file_path, sheet_name=sheet).assign(
        SALE_DATE=lambda x: fix_date_format(x['SALE_DATE'])
    ) for sheet in sheet_names], ignore_index=True)

print(f"Invalid dates (NaT): {sales_df['SALE_DATE'].isna().sum()} rows")
print(sales_df.head())
```

Listing C.1: Loading and Cleaning Excel Sheets

C.2 A.2 Inventory Optimization Model

```
import pandas as pd
import numpy as np
from scipy.stats import norm
from scipy.optimize import minimize

file_path = "/content/drive/MyDrive/Colab Notebooks/ABC-VEN/sales_data_fixed.csv"
df = pd.read_csv(file_path)
```

```

df['SALE_DATE'] = pd.to_datetime(df['SALE_DATE'])
df['MONTH'] = df['SALE_DATE'].dt.to_period('M')
df['YEAR'] = df['SALE_DATE'].dt.year

demand_data = df.groupby(['GROUP_PRODUCT', 'PRODUCT_NAME', 'MONTH'])['QTY'].sum().reset_index()
demand_summary = demand_data.groupby(['GROUP_PRODUCT', 'PRODUCT_NAME'])['QTY'].agg(['mean', 'std']).reset_index()
demand_summary.rename(columns={'mean': '\\mu D', 'std': '\\sigma D'}, inplace=True)

holding_cost_per_unit = 2
reorder_cost_per_order = 50
shortage_cost_per_unit = 5
overstock_cost_per_unit = 2
lead_time = 3
service_level = 0.95
Z = norm.ppf(service_level)

demand_summary['\\sigma D'].fillna(0, inplace=True)
demand_summary['SS'] = Z * demand_summary['\\sigma D'] * np.sqrt(lead_time)
demand_summary['ROP'] = (demand_summary['\\mu D'] * lead_time) + demand_summary['SS']

def total_cost(params, \\mu D, \\sigma D, SS, ROP):
    Q, I = params
    holding_cost = holding_cost_per_unit * I
    ordering_cost = reorder_cost_per_order / Q
    shortage_cost = shortage_cost_per_unit * max(0, (ROP - \\mu D * lead_time - I))
    overstock_cost = overstock_cost_per_unit * max(0, (I - ROP))
    return holding_cost + ordering_cost + shortage_cost + overstock_cost

```

```
optimal_results = []
for _, row in demand_summary.iterrows():
    \\mu D, \\sigma D, SS, ROP = row['\\mu D'], row['\\sigma D'], row['SS'], row['ROP']
    initial_guess = [max(1, \\mu D * lead_time), max(0, \\mu D * lead_time / 2)]
    result = minimize(total_cost, x0=initial_guess, args=(\\mu D, \\sigma D, SS, ROP),
                      bounds=[(1, None), (0, None)], method='L-BFGS-B')
    optimal_results.append({
        'GROUP_PRODUCT': row['GROUP_PRODUCT'],
        'PRODUCT_NAME': row['PRODUCT_NAME'],
        'Optimal Order Quantity': result.x[0],
        'Optimal Inventory Level': result.x[1],
        'Total Cost': result.fun
    })

optimal_df = pd.DataFrame(optimal_results)
final_results = pd.merge(demand_summary, optimal_df, on=['GROUP_PRODUCT', 'PRODUCT_NAME'])
print(final_results)
```

Listing C.2: Optimization with Cost Function

C.3 A.3 Machine Learning Model for ROP Prediction

```
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor, StackingRegressor
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
```

```python
import joblib

final_results = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/ABC-VEN/final_results.csv')
leakage_cols = ['SS', 'Optimal Order Quantity', 'Optimal Inventory Level', 'Total Cost', 'PRODUCT_NAME']
target = 'ROP'
features = [col for col in final_results.columns if col not in [target] + leakage_cols]

X = final_results[features]
y = final_results[target]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

numeric_features = ['\\mu D', '\\sigma D', 'STR']
categorical_features = ['GROUP_PRODUCT']
preprocessor = ColumnTransformer([
    ('num', StandardScaler(), numeric_features),
    ('cat', OneHotEncoder(handle_unknown='ignore'), categorical_features)
])

rf_pipeline = Pipeline([
    ('prep', preprocessor),
    ('rf', RandomForestRegressor(random_state=42))
])

rf_params = {'rf__n_estimators': [50], 'rf__max_depth': [3, 5], 'rf__min_samples_leaf': [20, 50]}
grid_rf = GridSearchCV(rf_pipeline, rf_params, cv=5, scoring='r2', return_train_score=True, n_jobs=-1)
grid_rf.fit(X_train, y_train)

best_rf = grid_rf.best_estimator_

```

```
stack_model = StackingRegressor(
    estimators=[(’rf’, best_rf)],
    final_estimator=LinearRegression(),
    cv=5
)

stack_model.fit(X_train, y_train)

# Evaluation
preds = stack_model.predict(X_test)
print(f"R2 Score: {r2_score(y_test, preds):.4f}")
joblib.dump(stack_model, "stacking_model.pkl")
```

Listing C.3: Stacking Ensemble for ROP Prediction