|  |
| --- |
| **COVENTRY**  UNIVERSITY |
| Faculty of Engineering, Environment and Computing |
| School of Computing, Electronics and Mathematics |
| MSC DATA SCIENCE (EECT109) |
| **7150CEM - Data Science Project** |
| DEVELOPMENT OF A MACHINE LEARNING BASED PREDICTIVE HEALTHCARE RESOURCE MANAGEMENT MODELS FOR PANDEMIC PREPAREDNESS AND RESPONSE |
| Author: BABATUNDE OBISANYA |
|  |
| SID: 13311980 |
| Supervisor: Dr. GANIYAT KAZEEM |
| Submitted in partial fulfilment of the requirements for the Degree of Master of Science in Data Science |
| Academic Year: 2023/24 |

**Declaration of Originality**

I declare that this project is all my own work and has not been copied in part or in whole from any other source except where duly acknowledged. As such, all use of previously published work (from books, journals, magazines, internet etc.) has been acknowledged by citation within the main report to an item in the References or Bibliography lists. I also agree that an electronic copy of this project may be stored and used for the purposes of plagiarism prevention and detection.

**Statement of copyright**

I acknowledge that the copyright of this project report, and any product developed as part of the project, belong to Coventry University. Support, including funding, is available to commercialise products and services developed by staff and students. Any revenue that is generated is split with the inventor/s of the product or service. For further information please see [www.coventry.ac.uk/ipr](http://www.coventry.ac.uk/ipr) or contact [ipr@coventry.ac.uk](mailto:ipr@coventry.ac.uk).

**Statement of ethical engagement**

I declare that a proposal for this project has been submitted to the Coventry University ethics monitoring website (https://ethics.coventry.ac.uk/) and that the application number is listed below (Note: Projects without an ethical application number will be rejected for marking)

Signed: Date:

Please complete all fields.

|  |  |
| --- | --- |
| First Name: | BABATUNDE |
| Last Name: | OBISANYA |
| Student ID number | 13311980 |
| Ethics Application Number | P165612 |
| 1st Supervisor Name | Dr Ganiat Kazeem |
| 2nd Supervisor Name | Dr Mark Johnston |

**This form must be completed, scanned and included with your project submission to Turnitin. Failure to append these declarations may result in your project being rejected for marking.**

# Abstract

In the face of increased challenges posed by COVID-19 and other pandemics in recent years, the development of machine learning models to predict healthcare resource management is paramount. Leveraging a robust healthcare dataset, this project aims to build upon the model proposed in a related work that used K-means clustering and support vector regression (K-SVR) for data clustering and prediction. The main objective of this project is to introduce another algorithm, random forest and compare its performance and accuracy with the algorithm used in earlier projects that target healthcare resource management.

The project unfolds with the compilation of a comprehensive public dataset from multiple regions, laying the foundation for the subsequent machine learning and time-series forecasting models. This dataset serves as the bedrock for training and building predictive models that are sensitive to regional variations, thus enhancing the adaptability and accuracy of the developed models.

The application of time series forecasting models is key to deal with the temporal dynamics of medical data. Predictive time series models and machine learning algorithms such as Prophet, Random Forest and Support Vector Regression were introduced. These models help predict future epidemics and hospitalizations and understand physiological patterns that affect the demand for health care resources.

An integral part of the project is the evaluation of the developed model. Various evaluation metrics are used to assess the efficiency and performance of each predictive algorithm. This rigorous evaluation process ensures that our predictive models are not only accurate but also reliable across a wide range of health conditions and geographic regions.

Through this comprehensive approach, the project strives to contribute to the field of predictive healthcare resource management. Extending the work of previous researchers, our introduction to Random Forests provides a comparative analysis and deepens the understanding of the strengths and weaknesses of various algorithms. Ultimately, the goal of this project is to provide healthcare professionals and policy makers with important ideas and tools that can facilitate decision-making in the area of pandemic preparedness and response.

# TABLE OF CONTENT

[Abstract 3](#_Toc13285)

[Acknowledgements 7](#_Toc18555)

[1 INTRODUCTION 8](#_Toc15910)

[1.1 Background of Study 8](#_Toc22736)

[1.2. Context and Applicability 9](#_Toc5123)

[1.3 Statement of Problem 9](#_Toc1603)

[1.4 Aim and Objectives of the Study 10](#_Toc13227)

[1.4.1. Intended Beneficiaries: 10](#_Toc4074)

[1.4.2. Research Questions 10](#_Toc15089)

[1.5 Scope of the Study 11](#_Toc29444)

[1.6 Limitation of the Study 11](#_Toc30217)

[1.7 Significance of the Study 11](#_Toc10030)

[2. LITERATURE REVIEW 12](#_Toc29008)

[2.1 Machine learning in Healthcare 12](#_Toc5545)

[2.2. Pandemic Preparedness and Response 13](#_Toc3601)

[2.3 Data-driven Approaches in Public Health 15](#_Toc27510)

[2.4. Healthcare resource management model. 16](#_Toc23200)

[2.5 Integration of Real-time Data in Predictive Models 17](#_Toc9607)

[3 Methodology 19](#_Toc12369)

[3.1 Introduction 19](#_Toc28514)

[3.2 Requirement 20](#_Toc28174)

[3.2.1 Dataset 20](#_Toc28643)

[3.2.2 Python 21](#_Toc7310)

[3.2.3 Jupyter Notebook 21](#_Toc15613)

[3.2.4 System Used 21](#_Toc29068)

[3.2.5 Machine Learning Regression Methods 22](#_Toc14004)

[3.3 Training and Testing of Machine Learning Models for Predictive Healthcare Resource Management 27](#_Toc4462)

[3.4 Modelling Strategy for Predictive Healthcare Resource Management A Scholarly Exploration 28](#_Toc19102)

[3.5 Hardware Requirements for Advanced Healthcare Predictive Modeling A Rigorous Examination 30](#_Toc11047)

[4 Data Analysis & Implementation 33](#_Toc27819)

[4.1 Introduction 33](#_Toc10356)

[4.2 Data Description 33](#_Toc14093)

[4.3 Design 34](#_Toc15509)

[4.4 Data Preprocessing 36](#_Toc29297)

[4.4.1 Correlation Heatmap 36](#_Toc11825)

[4.4.2 Data Cleaning 36](#_Toc4327)

[4.4.3 Requirement Elicitation Techniques 38](#_Toc23822)

[4.4.4 Strategic Analysis 39](#_Toc14640)

[4.5 Data Modelling 39](#_Toc19129)

[5. Result and Discussion 40](#_Toc11588)

[5.1 Introduction 40](#_Toc25950)

[5.2 Machine Learning Algorithms 40](#_Toc28244)

[5.2.1 Random Forest Regression (RFR) 40](#_Toc3669)

[5.2.2 Support Vector Regression (SVR) 43](#_Toc20886)

[5.2.3 Prophet Time Series Model 45](#_Toc25143)

[5.2.4 Over-fitted Results 47](#_Toc5267)

[5.3 Critical Evaluation 48](#_Toc13943)

[6. Project Management 50](#_Toc13525)

[6.1 Project Schedule 50](#_Toc8395)

[6.2 Risk Management 51](#_Toc11818)

[6.3 Quality Management 52](#_Toc16115)

[6.4 Social, Legal, Ethical and Professional Considerations 52](#_Toc7129)

[7 Critical Appraisal 54](#_Toc27034)

[8. Conclusion 55](#_Toc23658)

[8.1. Progress and Achievements 55](#_Toc4971)

[8.2 Students Reflection 56](#_Toc5383)

[9. Future Work 57](#_Toc23945)

[Bibliography and References 59](#_Toc6260)

**List of Figures**

[Figure 1 . Agile Methodology Phases 20](#_Toc31808)

[Figure 2 Structure of Random Forest (Chaya, 2020) 25](#_Toc9685)

[Figure 3 . Support Vector Regression structure. (www.educba.com) 27](#_Toc6973)

[Figure 4 . Design Activity Diagram 36](#_Toc5927)

[Figure 5 Correlation Heatmap 39](#_Toc7695)

[Figure 6 . Random Forest Scatter plot 43](#_Toc4055)

[Figure 7 . Random Forest Regression Density Plot 44](#_Toc24350)

[Figure 8 . Support Vector Regression Scatter plot 45](#_Toc25655)

[Figure 9 . Support Vector Density plot 46](#_Toc12528)

[Figure 10 . Prophet Time-series Model Scatter plot 47](#_Toc13199)

[Figure 11 . Prophet Time-series Model Density plot 47](#_Toc10920)

# List of Tables

[Table 1 . System Used 22](#_Toc6067)

[Table 2 : variables of the dataset. 34](#_Toc71)

[Table 3 . Random Forest Regression Evaluation Result 42](#_Toc26651)

[Table 4 . Support Vector Regression Evaluation Result 44](#_Toc10912)

[Table 5 : Over-fitted Support Vector Regression Result 48](#_Toc9136)

[Table 6 : Over-fitted Random Forest Regression Result 49](#_Toc3837)

[Table 7 : Over-fitted Prophet Model Result 49](#_Toc25879)

[Table 8 . Project Schedule 51](#_Toc23500)

[Table 9 : Risk Management 52](#_Toc14036)

# Acknowledgements

Embarking in this study adventure has been a transformative experience, and I enlarge my innermost gratitude to all the ones who've performed a role, each visible and unseen, inside the realization of this assignment. First and foremost, I explicitly express my honest appreciation to my supervisor, Dr Ganiat Kazeem, whose guidance, expertise, and unwavering assistance had been instrumental through each segment of this project. Her insightful feedback, encouragement, and determination to foster instructional excellence have been invaluable, shaping the trajectory of this research in profound ways.

I am thankful to my relatives for their unyielding assistance, understanding, and endurance for the duration of the ups and downs of this academic endeavor. Their encouragement furnished the emotional basis vital for me to persevere through demanding situations and celebrate achievement. To my colleagues and friends who shared their insights and perspectives, I appreciate you. Your discussions and collaborative spirit enriched the studies process, fostering a surroundings of shared getting-to-know and growth. The contributions of the school staff, from administrative assistance to library resources, can't go unnoticed. Their efforts behind the curtain had been critical in facilitating a conducive study surroundings. I also extend my appreciation to the anonymous individuals who contributed to the general public dataset that served as the spine of this research. Their dedication to open information has notably contributed to the development of studies in healthcare useful resource management.

Lastly, I need to acknowledge the broader academic community, both online and offline, for growing a surrounding of learning and exploration. The collective information and shared reviews inside this network had been a wellspring of inspiration. As I reflect on the finishing touch of this project, I am reminded that no scholarly undertaking is carried out in isolation. This assignment stands as a testament to the collaborative spirit and collective efforts of a numerous and supportive network. Thank you, one and all, for being vital elements of this academic endeavour.

# 1 INTRODUCTION

## Background of Study

With more than 3 million cases and 200,000 deaths at the end of April 2020 (WHO Corona-virus, 2023), the COVID-19 pandemic has rapidly emerged as a serious global health emergency (Fauci, et al 2020), testing the ability of healthcare systems to respond. The pandemic has created an unprecedented burden on healthcare facilities, requiring proactive management of essential resources and staff allocation. The burden on health systems stems from both the high infection rate of the pandemic and the fact that 20 to 30 percent of patients have moderate to severe disease/illness associated with multiple organ failure, prolonged periods of illness and hospitalization, and high mortality (CDC COVID-19, Vol. 2020). In addition, 5-12% of all patients with a diagnosis of COVID-19 and up to 33% of hospitalized patients require supportive intensive care in the intensive care unit (ICU) (Richardson, et al 2020). These estimates indicate that the number of patients transferred to intensive care for patients with COVID-19 is significantly higher than other inpatient ICU transfer rates of 11% (Moriarty, et al 2014).

In a parallel discourse, the comprehensive study by (Mohsin, et al. 2014), underscores the transformative impact of ML, particularly in adapting to the dynamic nature of pandemics. The authors discuss recent advancements in ML algorithms, highlighting their capability to not only analyse historical data but also to adapt in real-time, providing actionable insights for effective decision-making in healthcare resource allocation.

In the realm of workforce management, (Bajwa, et al. 2021) present recent advancements in ML methodologies, exploring their role in addressing staffing challenges in healthcare settings. The study discusses how ML models can leverage real-time data to anticipate staffing shortages, offering valuable insights for workforce optimization during the ongoing pandemic. The continuous refinement of ML strategies for workforce management is essential for ensuring the resilience of healthcare systems in the face of staffing uncertainties.

This study builds upon the foundational contributions of these researchers, aiming to develop an ML model tailored for predicting pandemic hospital resource utilization and staffing shortages in the current rapidly evolving landscape.

## 1.2. Context and Applicability

The application of machine learning to resource management in the healthcare sector is an escalating field that potentially reforms how healthcare authorities prepare for and respond to pandemics. As we go further into this project, it is important to comprehend the applicability and context of our research.

The ability of healthcare to efficiently allocate resources and effectively respond to pandemics directly influences public health. It improves the quality of patient care, the adaptability of non-stable conditions, and the decision-making processes healthcare authorities. To fully appreciate the significance of this research, it is essential to get a grasp of the healthcare landscape and tackle one of the problems that limit the use of machine learning for predictive modelling.

## 1.3 Statement of Problem

This research is based on the model proposed by (Tello, et al. 2022). The authors worked on a machine learning based forecast model for the prediction of inpatient bed demand. The K-means and Support vector regression (K-SVR) were used for clustering the data points and prediction. After a careful review of the work, we noticed that only SVR was used for prediction and the dataset was from a single healthcare facility that focused on inpatient bed demand. This procedure can limit the scope and areas in which the system can be used. To solve this problem of limitation, another regression algorithm will be implemented on a dataset from other healthcare facilities from multiple states/regions that cover both patients and staff data will be used.

## 1.4 Aim and Objectives of the Study

**Aim:**

The aim of this project is to develop a predictive model that anticipates health pandemic-related hospital resource demand (e.g., inpatient beds and ICU beds) and staffing shortages, helping healthcare facilities optimize their preparedness and response efforts.

**Objectives:**

1. Build and train machine Learning models using a dataset from multiple region.
2. Implement time series forecasting model to project future pandemic and hospital admissions, considering the temporal aspect of the dataset
3. Evaluate the machine learning and time-series models using different evaluation metrics.

### 1.4.1. Intended Beneficiaries:

This research is intended to assist, inform and support stakeholders in the healthcare sector

1. Healthcare Facilities: Hospitals and medical clinics directly benefit from effective resource management, which leads to positive improvements in patient care and outcomes.
2. Healthcare Authorities: Data-driven decisions during a pandemic will improve the preparedness and response of public health authorities and policy-makers.
3. Patients: The quality of care and the well-being of patients will benefit directly from timely and efficient resource allocation.
4. Society: In general, the population benefits from a stronger and adaptable healthcare system that is capable of managing unforeseen challenges such as pandemics.

### 1.4.2. Research Questions

1. Can Random Forest Regression perform better than the SVR for predicting Healthcare resource utilization?
2. Which data analysis techniques and machine learning tools are most suitable for this predictive modelling?

## 1.5 Scope of the Study

This study focuses specifically on the development of a machine learning model tailored for predicting pandemic hospital resource utilization and staffing shortages. The scope includes a comprehensive review of existing literature, analysis of disease trends, exploration of interdisciplinary approaches, and assessment of recent advancements in machine learning methodologies relevant to the study objectives.

## 1.6 Limitation of the Study

While the study aims to provide valuable insights, certain limitations are acknowledged. The model's effectiveness may be influenced by the availability and quality of data, and external factors such as policy changes or evolving medical practices.

## 1.7 Significance of the Study

This study's significance lies in its potential to contribute a specialized machine-learning model to the field of healthcare resource management during the pandemic. The developed model could enhance decision-making processes, optimize resource allocation, and mitigate staffing challenges, thereby improving the resilience of healthcare systems.

Making accurate decisions to predict healthcare resource needs during a pandemic has a tumbling effect on various aspects:

**Public Health:** Making predictions accurately has the potential to ensure that resources in the healthcare sector are available when and where they are needed. This results in positive outcomes for patients and public health in general.

**Response Planning:** Developing a predictive model contributes to informed decision-making for healthcare facilities, authorities, and policy-makers including resource allocation, staffing, and infrastructure planning.

**Decision-Making:** Decision-making based on previous data is gradually becoming crucial in healthcare. Predictive models aids healthcare facilities to optimize their procedures and make judgements that save lives.

**Adaptability:** Ability to adapt and manage dynamic situations is a hallmark of a strong global healthcare system. The insights resulting from our research project will contribute to the healthcare system’s resilience and adaptability

2. LITERATURE REVIEW

In this chapter, we review existing literature that serves as a foundation for this research. This review explores the field of healthcare resources management and predictive modelling during health crises, especially the COVID-19 pandemic. Gaining insight into different research work, we aim to understand strategies that have emerged over the years that manage hospital resource utilization and staff management. This chapter will present existing knowledge in this domain including methods, findings, and gaps in previous research work.

## 2.1 Machine learning in Healthcare

Tello et al. (2022) focused on constructing a machine learning-based forecast primarily addressing inpatient bed demand in the field of predictive modelling for healthcare resource management. The study focused on the essential issue of healthcare facility congestion, acknowledging its negative influence. The study’s major goal was to develop a machine learning technique for anticipating weekly inpatient bed demand, thereby aiding resource planning for the Emergency Department (ED) and Post-Anaesthesia Care Unit (PACU) and improving overall efficiency. The dataset used for this purpose included adult inpatient contacts at Geisinger Medical Center (GMC) over the previous five years. To build their machine learning strategy, the authors used the K-means clustering algorithm and the Support Vector Machine Regression technique (K-SVR). Notably, the K-SVR system demonstrated efficacy with a mean absolute percentage error (MAPE) ranging between 0.49 and 4.10% over the test period. The results also showed a reduction in variability, which translates to more stable outcomes after forecasting.

However, a critical examination of this work reveals certain drawbacks. This is because it is based entirely on GMC data, the research is notably focused on supporting only the GMC facility. Furthermore, the use of a single regression algorithm, while effective, may limit the findings’ generalizability to different healthcare settings. In order to solve these restrictions, our initiative aims to increase inclusivity and applicability in our project. Syed et al. (2023) contribute to the conversation by investigating the use of artificial intelligence (AI) in tailored treatment planning for congenital cardiac disorders. Their primary focus is on the prediction, risk stratification, and individualized therapy techniques for this particular medical problem. While useful for congenital cardiac problems, this work has a restriction in terms of disease-specific concentration, which may limit its generalizability to larger healthcare contexts. This highlights the need to take a balanced approach in our study, maintaining specificity in predictive modelling without sacrificing adaptability to various healthcare scenarios.

Shrikant et al. (2023) conducted an evaluation of machine learning techniques for COVID-19 case analysis, shifting the focus to the ongoing issues provided by the COVID-19 pandemic. Their investigation reveals the expanding landscape of machine learning in forecasting disease trends and patient outcomes in the COVID-19 era. Understanding how ML models adapt to changing patterns becomes critical for healthcare decision-makers as we face the difficulties of the current pandemic. This research provides important insights on the dynamic nature of ML applications, particularly in the setting of a quickly changing healthcare landscape. Junaid et al. (2021) add to the discussion by investigating the interface of machine learning and epidemiology, stressing their synergies in delivering unique insights into healthcare emergencies. Their focus on unravelling complicated patterns and providing techniques for maximizing resource allocation during pandemics is consistent with our project’s broader goals. The interdisciplinary approach described in their work highlights the potential for machine learning to greatly contribute to understanding and addressing healthcare crises. Finally, (Fu-Yuan Cheng, Himanshu Joshi, and Pranai Tandon’s 2020) work focuses on utilizing ML to predict ICU transfers in COVID-19 hospitalized patients. While useful for ICU transfers in COVID-19 patients, this disease-specific application creates a void in understanding the broader uses of machine learning in predictive healthcare resource management. It emphasizes the importance of taking a nuanced approach in our endeavour, balancing disease-specific concerns with general applicability.

In summary, the literature analysis sheds light on the expanding landscape of machine learning applications in healthcare resource management. It emphasizes the significance of resolving restrictions such as disease-specific foci, dataset inclusivity, and algorithm variety. Based on these findings, the project will contribute to the creation of more universally applicable and robust predictive models, ensuring their relevance and usefulness in a variety of healthcare contexts.

## 2.2. Pandemic Preparedness and Response

Karaarslan, E., & Aydın, D. (2021) used rigorous methodologies with a focus on ward-level granularity to anticipate ward-level bed requirements for pandemic resource planning. While their approach was successful in making predictions at this level, the study lacked an in-depth examination of scalability and external validation, giving a significant lesson but leaving important elements unaddressed. Notably, the dataset used is not given, and the research may benefit from a more thorough examination of model comparison measures. Johnson, et al. (2023) add to the conversation by developing an AI-based decision support and resource management system for the COVID-19 pandemic. The paper’s strength is its emphasis on AI and decision assistance, however it lacks detailed model evaluation measures, making it difficult to assess the robustness of their method. While the study was useful in providing decision assistance during the pandemic, it fell short in giving a detailed discussion of evaluation measures, leaving space for improvement in this key area. Keshavamurthy, et al. (2022) undertook a systematic assessment of machine learning and deep learning algorithms for infectious disease prediction, providing a thorough overview. The research, however, may benefit from a more in-depth examination of the real-world implementation issues connected with these approaches. The lack of a specific dataset, as one would anticipate from a review study, emphasizes the need for more research that bridges the gap between theoretical efficacy and practical obstacles in real-world implementation. Cheng, et al. (2020) demonstrated success in a crucial environment by applying machine learning to predict ICU transfers in hospitalized COVID-19 patients. The study, however, falls short of exploring the model’s generalizability to non-COVID-19 settings, leaving a crucial component unexplained. The lack of detailed details regarding the dataset utilized, as well as the emphasis on generalizability, make comprehending the prediction model’s broader applicability difficult. Myers, et al. (2020) investigated the features of COVID-19-infected hospitalized patients, offering real-world patient insights. The study descriptive nature, on the other hand, limits its usefulness to predictive modelling for resource management. The emphasis on patient characteristics rather than predictive models for resource management highlights a gap in addressing the special needs of pandemic healthcare resource optimization. These authors emphasize the importance of scalable, well-evaluated predictive models in pandemic healthcare resource management. The observed gaps, such as minimal examination of external validation, scalability, and precise assessment metrics, guide our project’s target areas for constructing a viable predictive model. These deficiencies are addressed by the research topics, which attempt to compare model performance and identify the best data analysis approaches and machine learning tools for successful predictive modeling in healthcare resource management during pandemics.

## 2.3 Data-driven Approaches in Public Health

Tello et al. (2022) tackled the significant issue of healthcare facility overcrowding by developing a Machine Learning-based prediction to predict inpatient bed demand. The study, which was conducted at Geisinger Medical Centre (GMC), aimed to maximize resource planning for the Emergency Department (ED) and Post-Anaesthesia Care Unit (PACU). The study drawback is its facility-specific methodology and reliance on a single regression algorithm, despite demonstrating the usefulness of the K-SVR system in minimizing variability. Our effort intends to address this constraint by incorporating a variety of datasets and applying other regression algorithms for broader applicability.To combat the COVID-19 epidemic and future public health concerns, Ros et al. (2021) pushed for a worldwide, data-driven systems strategy. While emphasizing the value of international collaboration, the research is vague about the datasets used and does not go thoroughly into potential ethical implications. The emphasis on global cooperation is notable, but a more nuanced examination of the problems connected with such collaboration is required. Karaarslan and Aydn (2021) contribute a COVID-19 pandemic AI-based decision support and resource management system, demonstrating the potential of artificial intelligence in healthcare decision-making. However, the research falls short of discussing broader public health contexts beyond COVID-19, and the lack of particular dataset and scaling specifics limits the suggested system’s application to different public health concerns. Keshavamurthy et al. (2022) give a systematic study of machine learning and deep learning algorithms for predicting outcomes.

While the paper provides a comprehensive overview of methodologies, it lacks a detailed analysis of challenges encountered during real-world implementation. Due to the lack of a specific dataset for the review, insights into the practical applicability of the discovered techniques to various public health scenarios are limited. The writers all underlined the potential of data-driven techniques in public health, but they also had the same shortcomings, such as a lack of specificity in datasets, insufficient discussion of implementation issues, and a focus on individual diseases rather than a broader public health context. The project will look to fill these gaps by conducting a comprehensive examination of data-driven methodologies in various public health scenarios, addressing implementation issues, and ensuring scalability and adaptability beyond specific diseases. As a result, the research topics are aligned with the identified gaps, with the goal of developing a comprehensive understanding of data-driven approaches in public health and their practical implications for pandemic preparedness.

## 2.4. Healthcare resource management model.

Tello et al. (2022) explored healthcare resource management by creating a Machine Learning-based forecast for inpatient bed demand. While the study adequately addresses the issue of healthcare facility overcrowding, notably at the Geisinger Medical Center (GMC), it falls short of offering a comprehensive analysis of alternate models and lacks a broader context outside the individual hospital. The K-SVR system’s strength is in reducing variability, however the sole focus on GMC limits the suggested model’s generalizability to varied healthcare contexts. Votto et al. (2021) carried out a thorough evaluation of the literature on the function of artificial intelligence in tactical human resource management. The study provides a broad overview of AI applications, although its relevance to healthcare is limited. The strength resides in the ability to synthesize AI applications in human resource contexts; nevertheless, the lack of focus on healthcare settings restricts its direct relevance to our theme. This reveals a void in the research about AI models specifically designed for healthcare resource management. Sahoo and Goswami (2023) present an in-depth examination of Multiple Criteria Decision-Making (MCDM) approaches, providing insights into decision-making advancements. While the review is comprehensive and useful, its focus goes beyond healthcare resource management, making it difficult to select models that are directly related to our issue. The work succeeds at describing advances in decision-making procedures but falls short in healthcare contexts, underlining the need for literature that tailors MCDM methods to healthcare resource optimization. Karaarslan and Aydn (2021) describe a COVID-19 pandemic decision support and resource management system based on artificial intelligence. The study demonstrates the potential of AI in healthcare decision-making, however it focuses exclusively on the COVID-19 setting. The model’s strength is its relevance to pandemic scenarios, but its applicability to non-pandemic contexts is limited due to a lack of specificity on the broader healthcare resource management model. The study emphasizes the importance of adaptive AI models that can address a variety of hospital resource management concerns. All of the literature studied demonstrates a variety of models and approaches to hospital resource management. However, common drawbacks include a lack of specialization in the context of healthcare, a concentration on specific institutions, or a general lack of human resource management. The project intends to fill these gaps by constructing adaptive models that take into account the larger healthcare resource management landscape, which includes a wide range of healthcare settings and difficulties. The research topics are aligned with these highlighted deficiencies, with the goal of developing models that are versatile and effective in a variety of hospital resource management settings.

## 2.5 Integration of Real-time Data in Predictive Models

Mukherjee et al. (2022) concentrated on integrating real-time data for inpatient bed demand forecast, using a K-SVR method to reduce variability. The study’s downside, while admirable, is its facility-specific approach to Geisinger Medical Center (GMC) and reliance on a single regression model, which limits generalizability. Our effort tries to fill these gaps by combining different datasets and regression techniques to create a more adaptive predictive model. Machine learning was used by Saadatmand et al. (2023) to predict ICU admission, mortality, and length of stay in early-stage COVID-19 patient admission. While leveraging real-time data successfully, the study’s emphasis on COVID-19 patients limits its broader usefulness. Furthermore, the study lacks context on the datasets employed, preventing a thorough understanding of the model’s generalizability. Our initiative aims to go beyond specific diseases by using a variety of real-time datasets to create a more robust forecasting model. Mukherjee et al. (2021) investigated the Internet of Health Things (IoHT) for personalized healthcare through the integration of edge-fog-cloud networks. While emphasizing the possible benefits, the report is vague about the challenges of real-time data integration, and the datasets used are not specifically described. This literature emphasizes the importance of conducting a rigorous assessment of real-time data integration difficulties and datasets, guiding our project’s focus on extensive exploration and tackling real-time data integration challenges. The concept of a digital twin for intelligent context-aware IoT healthcare systems was suggested by Elayan et al. (2021). The study stresses individualized healthcare but does not go into detail about the challenges of real-time data. This indicates a gap in the literature, leading our initiative to address real-time data difficulties in healthcare predictive models for better informed and effective resource management. These authors emphasized the significance of incorporating real-time data into prediction models for healthcare resource management. However, prominent disadvantages include a lack of generalizability, dataset specialization, and insufficient examination of real-time data integration issues. The project’s goal is to close these gaps by creating a more adaptive and comprehensive prediction model, leveraging varied datasets, and tackling issues related with real-time data integration in hospital resource management. The research topics are aligned with these highlighted deficiencies, with the goal of developing a versatile, educated, and effective predictive model in a variety of healthcare contexts

# 3 Methodology

## 3.1 Introduction

In this project, developing predictive healthcare resource management models based on machine learning, the Agile methodology stands out as an adaptive and focused approach to project management. The phases of the agile methodology consist of Analysis, Plan, Design, Build, Test, Review, Launch and the process goes over again till the end of the product life cycle (William Goddard, 2021). In the context of health and complexity, dynamic environments characterized by growth requirements and unexpected challenges, the principles of Agile methodology fit perfectly with the need for constant adaptation and robust development. Agile and iterative changes, also known as sprints, are suitable for the healthcare sector and can create, test and deliver functional units in short and flexible periods. This flexibility is essential in healthcare, where data and priorities change rapidly. Agile means that predictive models can be continuously improved to meet new requirements and challenges in the healthcare industry.



Figure 1. Agile Methodology Phases

(William Goddard, 2021)

Agile adaptability is important when dealing with uncertainty in medical data. This methodology and its commitment to flexibility allows predictive models to adapt perfectly to the constantly evolving healthcare environment requirement, incorporating unexpected changes in patient characteristics, morbidity and treatment methods. Additionally, Agile methodology's emphasis on collaboration and communication are important to the development of predictive models. Revolutionary changes are giving more room for closer interaction between cross-functional teams, data scientists and healthcare professionals. In the context of predictive healthcare resource management, this collaborative approach ensures that models not only meet technical standards but are effective in addressing the unique challenges healthcare professionals face. As shown in studies by (Dyba, Dingsoyr and Moe 2014) and (Beck et al. 2001), agile methods and customer-oriented approaches involve stakeholders throughout the development process will provide valuable feedback. This process will ensure that predictive models are not only technical, but also aligned with the needs and expectations of healthcare workers. In conclusion, the Agile methodology emerges as a strategic option and good support for the development of predictive models for health resource management. An adaptation, transformation and collaboration framework is a useful way to explore the complexity of health systems and build predictive models that are relevant and adaptable to a changing environment.

## 3.2 Requirement

The technology, methods, and resources needed to effectively complete this project will be discussed in this section. We want to use different technologies like machine learning and tableau visualization to predict global healthcare availability and quality. We will also go through the framework used to create the system and make accurate predictions.

### 3.2.1 Dataset

The dataset for this research work is a Time-series dataset that was obtained from a public website was used for this project. Time restraints, the nature of the project, and the title all played a part in the decision to employ a secondary dataset. The (catalog.data.gov) provided the healthcare resource management dataset that was utilized. Due to its applicability for research and relevance to the project's goals, this dataset was chosen.

### 3.2.2 Python

Python is a widely used and flexible programming language. It has many different applications in industrial sectors. It is often used for tasks such as data analysis, web development, machine learning, prototyping, and testing software. Python and its growing popularity lies in its extensive library ecosystem and easy-to-understand syntax. Attract professionals with different backgrounds (Coursera, 2023). Python in particular has powerful modules like NumPy, Pandas, and scikit that teach it perfectly to simplify the handling, processing, and analysis of large amounts of data. These modules also provide advanced tools to implement complex machine-learning algorithms, enabling tasks such as classification, regression, clustering, and recommendation systems.

### 3.2.3 Jupyter Notebook

One of the major positive perspective of utilizing the Jupyter Notebook is its capacity to combine coding and documentation on a single page. The scratch pad arrange permits us to embed code cells and informative content, making it an incredible stage for composing each step of the machine learning show improvement preparation. This integration progresses extend conveyance and advances collaboration as group individuals can effortlessly get it and contribute to the code base.

### 3.2.4 System Used

To achieve this project’s computational requirement, the specification of the computer used for this project are as follow

Table 1. System Used

|  |  |
| --- | --- |
| **Specification** | **Details** |
| Type | SGIN |
| Processor | Intel(R) Celeron(R) N4500 @ 1.10GHz 1.11 GHz |
| System type | 64-bit operating system, x64-based processor |
| Memory | 12.0 GB (11.8 GB usable) |
| Windows OS | Windows 11 Home |

### 3.2.5 Machine Learning Regression Methods

Artificial intelligence (AI) and machine learning are making rapid advances in healthcare and can improve disease self-management, treatment and well-being. Machine learning can be used to discover hidden patterns and relationships to predict outcomes and create intelligent healthcare systems (Panesar, 2021). Because these technologies offer the ability to collect and monitor patient health and lifestyle characteristics, the use of smart-phones, wearable, and the Internet of Things (IoT) has accelerated the shift in healthcare from volume-based care worldwide

Regression methods based on machine learning have become useful tools for assessing and evaluating the quality of health services worldwide. With the help of these algorithms, we can not only predict health outcomes, but also identify the factors that influence access and quality of healthcare. These algorithms use statistical methods to compare relationships between various input data and target variables. This section covers the use of machine learning regression approaches to assess healthcare management and overall quality, as well as their advantages, disadvantages, and future directions. The availability and quality of global health can be assessed using machine learning regression techniques such as random forest regression, support vector regression, etc. These technologies allow predictive modelling, the identification of variables related to the quality of care, the investigation of spatial variations and the management of resources.

The Prophet Time series model is a powerful and robust tool for the management of health resources and staff allocation. It is ideal for using historical patient data to predict future trends during infectious disease outbreaks or pandemics. By focusing on capturing seasonal and daily trends, Prophet helps health systems allocate resources to prevent mismanagement and short staffing. This helps prepare for possible increases and increase the efficiency of health resource management during critical periods (Qureshi et al., 2020).

**Random Forest Regression Algorithm**

Random Forest is a general algorithm built from decision trees mainly for classification and regression problems. This algorithm is developed by combining several decision trees and can be thought of as a forest consisting of a combination of these trees. Each decision tree is trained using different subsets of data and a random selection of features. It creates many trees from different perspectives to store additional attributes and relationships in the data. The purpose of Random Forest is to obtain more reliable and stable predictions by combining the predictions of each tree. In addition, it reduces overhead and is efficient in processing large-scale data. Random Forest can also be used to evaluate features such as feature selection and prioritization and is often used in industry and academia (Atala et al. 2023). The steps below presents an overview of the RF model step by step process.

**Bootstrapped Sampling:** The algorithm starts by creating several subsets of the original data set using bootstrap sampling. This requires selecting random data points with new sets of replacements for each tree in the forest.

**Random selection of features:** For each decision tree, a random subset of the entire set of features is selected. This increases the diversity of the trees by preventing them from relying on the same characteristics and improves the model and its ability to capture a wide variety of patterns.

**Decision Tree Training:** Each random forest tree is trained independently based on each subset of the data. During training, the tree makes decisions based on selected features, recursively dividing the data into subsets until it reaches a certain depth or a predetermined stopping criterion.

**Averaging (regression):** After all the trees are trained, they together contribute to the final prediction. In classification tasks, the class that receives the most votes from individual trees is chosen as the predicted class. Regression tasks take the average of all tree predictions.

**Ensemble Output:** The final output is the combined output of all the individual trees. This tool helps reduce overfitting and reduces the effect of noise on the data eventually resulting in a more robust and accurate model.

Random Forest Regression is a powerful predictive modeling technique used to predict the availability and quality of health services. using a bagging technique and generating multiple decision trees from randomly selected subsets reduces overheating and improves generalization properties. The ability to analyze large data sets and identify complex relationships provides insights into healthcare. The method and its ability to handle high-dimensional data, capture non-linear correlation, and provide information about future significance make it a valuable tool for building accurate regression forecasts with applications in various health fields such as finance, medical research and ecological research.

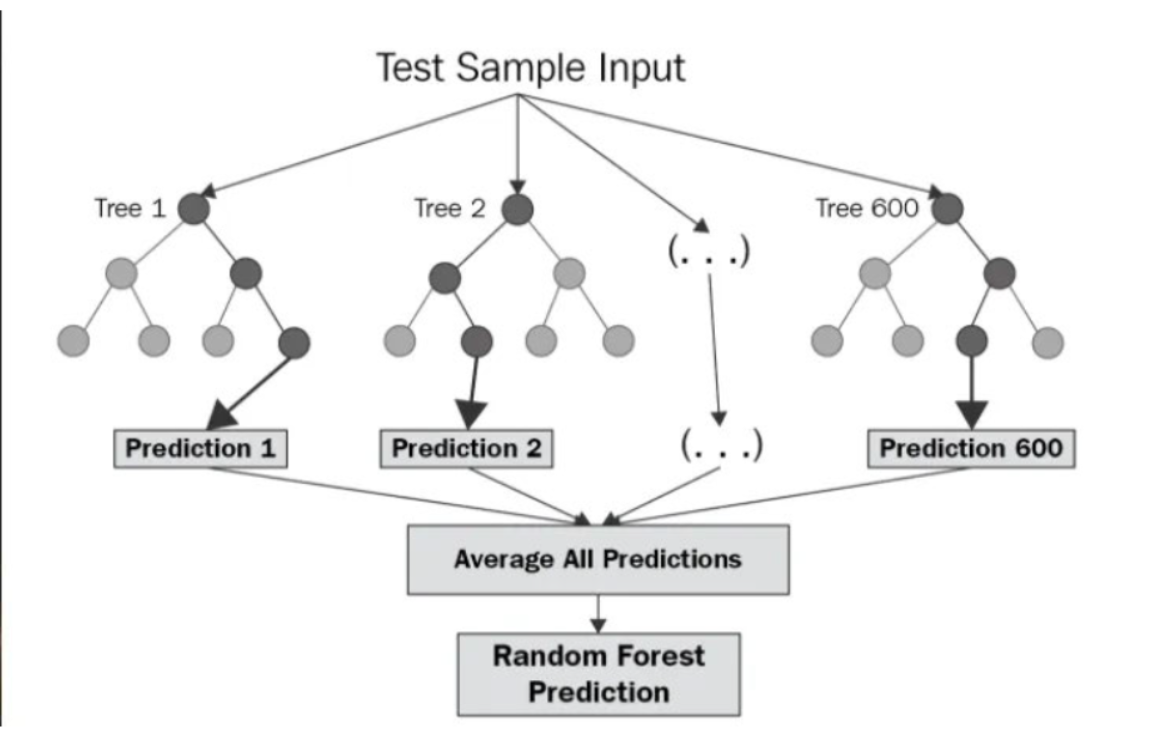


Figure 2 Structure of Random Forest (Chaya, 2020)

The structure of the random forest is shown above. The parallelism of trees and their complete lack of interaction is obvious. During training, RF builds multiple decision trees and outputs the class average as a prediction of all trees.

**Support Vector Regression Algorithm**

Support Vectors are systems which use hypothesis space of a linear function in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory. Support vectors was initially popular with the NIPS community and now is an active part of the machine learning research around the world. SVR becomes famous when, using pixel maps as input; it gives accuracy comparable to sophisticated neural networks with elaborated features in a handwriting recognition and prediction tasks (Andrew, 2020). It is also being used for many applications, such as predictive modelling, face analysis, and so forth, especially for regression-pattern and classification based applications.

SVR is a supervised learning algorithm that is used for tasks related to regression. one of the aims of the SVR is to find predicting functions that continue to generate outcomes while reducing errors in prediction. Below is the step by step of how it works.

**Data Representation:** SVR represents data points in a high-dimensional space and attempts to find a hyperplane that best fits the data.

**Feature Transformation:** Non-linear features are transformed into a higher-dimensional space using a kernel function. This allows SVR to capture complex relationships between variables.

**Margin Creation:** SVR introduces a margin around the predicted hyperplane, where data points are allowed to fall without penalty. The width of this margin is determined by a parameter called ε (epsilon).

**Loss Function:** SVR uses a loss function that penalizes points falling outside the margin. The loss increases with the distance of a data point from the predicted hyperplane.

**Hyper parameter Tuning:** SVR involves tuning hyperparameters, such as the choice of kernel (linear, polynomial, or radial basis function), regularization parameter (C), and kernel-specific parameters.

**Optimization:** The goal of SVR is to find the hyperplane that minimizes the loss function and, consequently, prediction errors.

**Prediction:** Once the model is trained, it can be used to make predictions on new, unseen data.

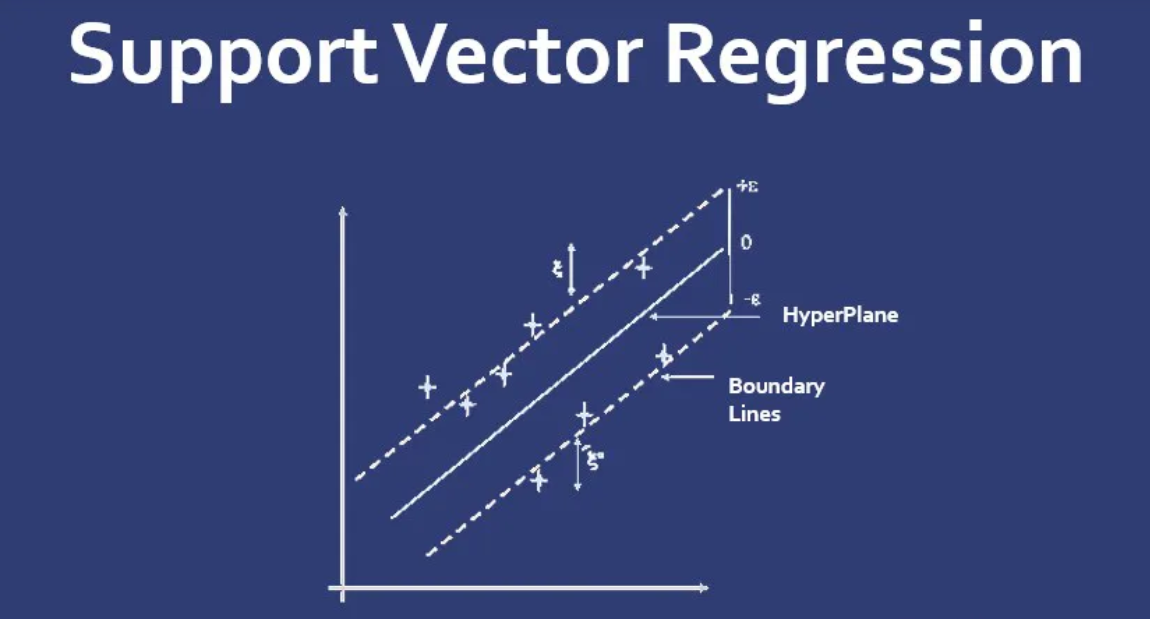


Figure 3. Support Vector Regression structure. ([www.educba.com](http://www.educba.com))

Below sheds light on the figure above.

1. Hyperplane: It is a dividing line between two classes of data in a higher dimension than the real dimension. In SVR, it is defined as a line that helps predict the target value.

2. Kernel: In SVR, regression is performed in a higher dimension. For that we need a function that should map the data points to their higher dimension. This function is called kernel. The kernel used in SVR is Sigmoidal Kernel, Polynomial Kernel, Gaussian Kernel etc.

3. Boundary lines: These are two lines drawn around the hyperplane at a distance of ε (epsilon). It is used to create a margin between data points.

4. Support vector: This is the vector used to define the hyperplane or we can say that these are the extreme data points of the data set which help to define the hyperplane. These data points are located near the boundary.

## 3.3 Training and Testing of Machine Learning Models for Predictive Healthcare Resource Management

The training and testing phase of constructing machine learning models for predictive healthcare resource management is crucial. This section expands on the approaches adopted and defends these choices using existing literature.

**I Model Training**

1. Choosing a Dataset

In the context of the study design and topic, the data sample selected represents a subset of the entire data set that contains important features for predicting health resource use during a pandemic. The focus is on variables such as critical staff shortages, availability of hospital beds and occupancy of adult intensive care units, which are key factors affecting the total number of adult patients hospitalized for confirmed cases of COVID-19. This selection method is suitable for the research topic and hypothesis because it ensures the inclusion of variables crucial to predicting and understanding the health resource effects of the pandemic.

2. Feature Engineering As described in (Fan et al., 2019), feature engineering is critical to model performance. The model's performance is improved by extracting relevant characteristics from the dataset, such as patient demographics, infection rates, and past resource utilization.

3. Literature on Algorithm Selection (Khetani et al., 2023) evaluates the performance of several algorithms in healthcare predictive modelling, emphasizing the importance of algorithm selection. The algorithms used, such as Random Forest Regression, Prophet model and Support Vector Regression, were chosen based on their documented efficacy in hospital resource management applications, ensuring that the models aligned with the research aims.

**II. Model Testing**

1. Dataset Segmentation

(Soper, 2021) emphasizes the significance of suitable dataset separation for training and testing. Using approaches such as k-fold cross-validation ensures robust model evaluation, avoiding over-fitting and offering a more realistic picture of model performance.

2. Metrics for Evaluation

Hodson (2022) highlights the use of appropriate evaluation measures for assessing prediction model accuracy, such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The testing phase allows a quantifiable and comparable assessment of the models' predictive capabilities by applying defined evaluation metrics, complying with best practices in healthcare predictive modelling.

3. The Interpretability of the Model

(ElShawi et al., 2021) discusses the interpretability of machine learning models in healthcare, emphasizing the significance of comprehending model predictions. Ensuring model interpretability during testing is critical for generating increased acceptability and utilization by allowing healthcare practitioners to trust and comprehend the predictions.

The approaches used in the project to train and evaluate machine learning models are based on best practices documented in the literature. This approach ensures the reliability, generalizability, and interpretability of the generated predictive healthcare resource management models by leveraging insights from prior studies.

## 3.4 Modelling Strategy for Predictive Healthcare Resource Management A Scholarly Exploration

The modeling part of the study technique and methodology is critical to the project’s success. This section thoroughly outlines and justifies the chosen modeling methodologies using current research as a foundation.

**II. Algorithm Selection**

1. SVR vs. Random Forest Regression

The application of Random Forest Regression and Support Vector Regression (SVR) in healthcare resource prediction has been thoroughly studied in comparative research (He et al., 2020; Srivatsan et al., 2020).

Random Forest Regression is recommended in healthcare situations with multifarious variables because of its ability to manage complex interactions and offer robust predictions (Delcea et al., 2023).

SVR, which is known for capturing non-linear patterns, provides an alternative, allowing for a thorough comparison of their performance in predicting healthcare resource usage (Wang et al., 2023).

Created and open sourced by Facebook, the Prophet model is a powerful tool specifically designed for time series forecasting. Unlike traditional forecasting models, Prophet is tailored to deal with time-series data issues such as seasonality, holidays and outliers, making it particularly well-suited for forecasting healthcare resource use over time.

2. Ensemble Learning Methods

Ensemble learning strategies, such as those described in (Ganaie et al., 2022), have been shown to improve model accuracy and generalization. Using ensemble techniques, such as aggregating predictions from many models, improves the resilience and reliability of predictive models, matching with machine learning best practices (Cao et al., 2020).

**III. Data Analysis Methods and Machine Learning Software**

1. Analysis of Feature Importance

The outcomes of the study (Gupta et al. 2023) emphasize the necessity of feature importance analysis in healthcare predictive modelling. Using approaches such as SHAP (Shapley Additive exPlanations) values enables for the analysis of the impact of various variables.

2. Hyper-parameter Adjustment

The tweaking of hyper-parameters, as described in (Schratz et al., 2019), is critical for T model performance. Using approaches like grid search or Bayesian optimization ensures that the models are fine-tuned to get the best possible outcomes, contributing to predictive modelling overall performance.

Model Validation and Evaluation

1. Techniques for Cross-Validation

K-fold cross-validation, as discussed in (Said et al., 2020), is a widely used technique for evaluating robust models. Using k-fold cross-validation ensures an unbiased evaluation of model performance, reducing over-fitting and offering a more realistic estimate of predictive capabilities.

2. Ethical Considerations

As highlighted in (Johnson, 2019), ethical considerations in machine learning underline the significance of fair and unbiased models in healthcare. Adhering to ethical criteria during the modelling phase is crucial to avoiding exacerbating current healthcare inequities and ensuring responsible predictive model deployment. As indicated by the available research, the selection of modelling methodologies is based on a scholarly understanding of machine learning applications in healthcare. The modelling approach aligns with the research questions by combining established algorithms, leveraging ensemble learning, and incorporating ethical considerations, addressing the comparative performance of Random Forest Regression and SVR while ensuring the reliability and fairness of predictive healthcare resource management models.

# 4 Data Analysis & Implementation

## 4.1 Introduction

The method of gathering, purifying, classifying, and processing raw data that allows you to derive pertinent and useful records for companies is called data analysis. (Karin Kelley,2023). Data analysis is a radical method that includes gathering, purifying, organizing, and studying unstructured information that allows you to produce insightful findings and actionable information for companies. It consists of a number of approaches, tools, and strategies for recognizing patterns, trends, and connections in information. Data analysis overarching cause is to empower organizations to make reasoned decisions, boost operational effectiveness, spot opportunities, and advantage and benefit of their respective markets

## 4.2 Data Description

The dataset used for this research project was gotten from Data.gov website. The dataset with 109 features and 69,993 rows, focusing on relevant features. The dataset is selected to minimize data acquisition costs and align with the project's time constraints. According to the publishers, the dataset is been updated on a monthly basis to maintain authenticity as new data is generated. The dataset provides state-aggregated data for hospital utilization in a time-series format dating back to January 1, 2020. These are derived from reports with facility-level granularity across three main sources: (1) HHS TeleTracking, (2) reporting provided directly to HHS Protected by state/territorial health departments on behalf of their healthcare facilities and (3) National Healthcare Safety Network before July 15, 2020. (Data.gov, 2020). The table below presents some of the features of the dataset.

Table 2: variables of the dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| No | Variables | Data Types | Descriptions |
| 1 | total\_staffed\_pediatric\_icu\_beds total\_staffed\_pediatric\_icu\_beds\_coverage | Integer | Number of beds allocated to staffs in ICU |
| 2 | State | String | Acronym for each state in the US |
| 3 | Date | Date | Date for each entry |
| 4 | critical\_staffing\_shortage\_today\_not\_reported | Integer | How much staffs shortage was recorded |
| 5 | staffed\_icu\_adult\_patients\_confirmed\_covid | Integer | List of staffs that became a covid-19 patient |
| 6 | adult\_icu\_bed\_utilization | Integer | Number of beds per patient. |
| 7 | deaths\_covid | Integer | Number of deaths per day in each state |
| 8 | previous\_day\_deaths\_influenza\_coverage | Integer | Number of death in previous day |
| 9 | previous\_day\_admission\_adult\_covid\_confirmed | Integer | Number of patients taken indaily |
| 10 | adult\_icu\_bed\_utilization\_numerator | Integer | Number of beds used in icu. |
| 11 | adult\_icu\_bed\_utilization | Float | Average beds used per icu rooms |

## 4.3 Design

In this research the design method being applied is the top-down design approach. this method is used in various computing Fields. it is used in homogeneous systems and corresponds a number of steps in the building of a distributed diagrammatic representation of the project starting from scratch.

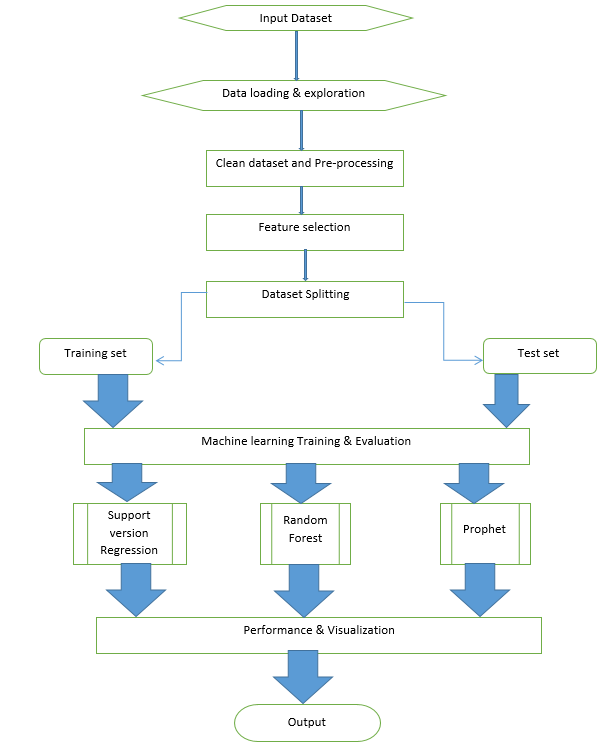


Figure 4. Design Activity Diagram

In the above design flowchart, a step-by-step top-down machine learning system for predicting healthcare ML model for predicting resource utilization is clearly detailed about the progress of the system. The system performs the entire process of machine learning, Data input and preprocessing, research/analysis, data cleaning until the result/output is obtained.

## 4.4 Data Preprocessing

Data preprocessing, which involves cleaning, transforming and integrating raw data to make it ready for further analysis, is an important step in data analysis and data mining. Its main purpose is to improve data quality and suitability for a specific data mining activity. In this step, various methods are used to handle inconsistent or missing data, remove duplicates, and correct errors. Data can also be modified to create new features and scale or normalize values. Data from multiple sources must be integrated into a single dataset. Researchers can ensure that the data used in the analysis are correct, complete and suitable for the chosen data mining method by performing extensive data pre-processing, which ultimately leads to more reliable results.

### 4.4.1 Correlation Heatmap

A correlation heatmap is a visual representation of the relationship between two different dimensions in the form of a colour table. Data are presented on a colour scale of different colour-coded cells. The values of the first dimension are displayed as rows, while the values of the second dimension are displayed as columns. The colour intensity of each cell range reflects the frequency or number of measurements corresponding to a certain set of measurement values. Data analysis can benefit from correlation heatmaps because they visually show trends, differences, and variance in data. They are enhanced with a colour bar to help interpret and understand the information displayed on the heatmap (Kagoyal, 2020). Figure 5 presents the heatmap for the first 37 attributes of the dataset.

### 4.4.2 Data Cleaning

In the data preprocessing phase, data cleaning is an important step that aims to improve and correct a dataset quality. The process entails spotting and fixing many kinds of incomplete data including blank cells, information in the wrong format, and errors in entries. Empty cells can create errors and inhibit analysis of data, and improperly formatted data might make analysis difficult to do precise computations comparisons. The integrity and reliability of the used dataset can be greatly impacted by inaccurate or inconsistent data such as outliers or wrong numbers. Various methods can be used to clean the data. For instance, extra features can be dropped to simplify the layout and draw attention to the most important information. a data frame's index can be switched to improve the data structure and organization. A CSV file can be made more efficient by removing rows that are not needed and by renaming columns with labels that are easier to understand. Collectively, these methods help to increase the dataset's overall quality and usability. In this research after evaluating the dataset, null values were checked for which was found. The index was changed to the third feature when generating the heatmap, The null values were replaced with zero to improve the Dataset usability. Also a values in some columns were changed from object to integer, and some features were dropped as they did not have a strong correlation from the heatmap generated.

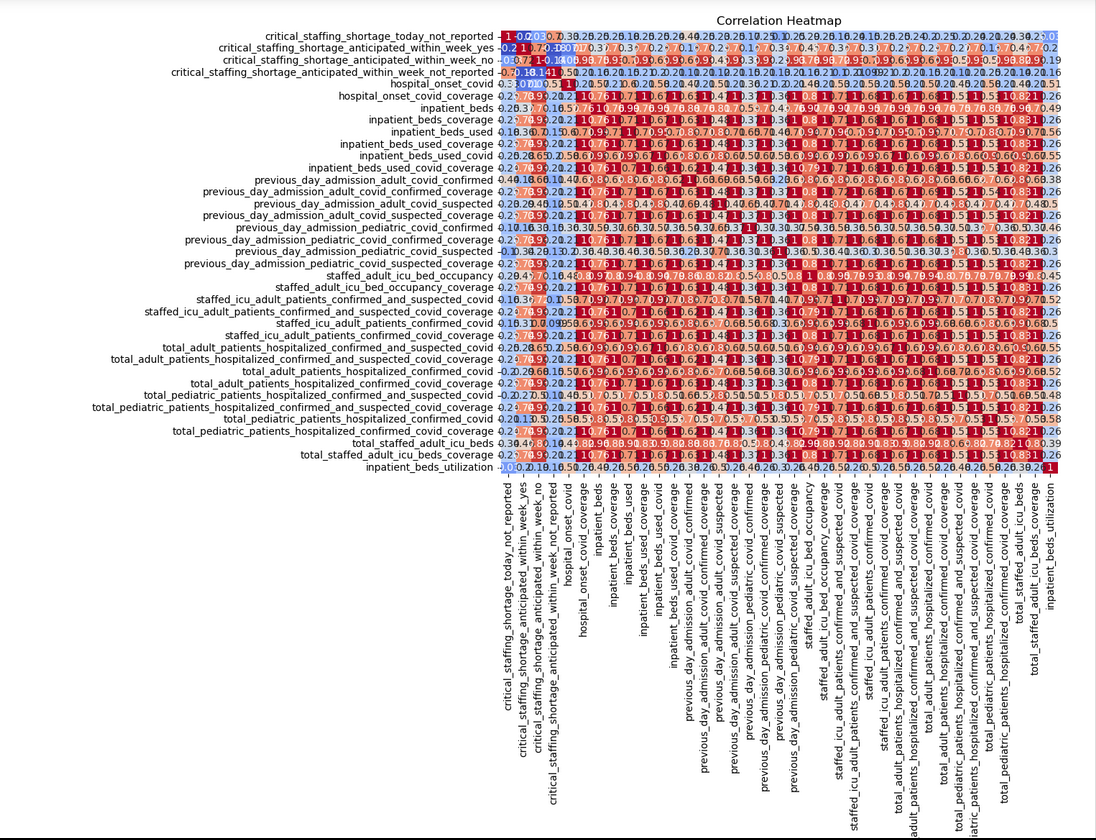


Figure 5 Correlation Heatmap

### 4.4.3 Requirement Elicitation Techniques

The data collection process combines several methods to ensure a comprehensive understanding of the functional and non-functional sectors. Consumer interviews, including health professionals, and general user surveys were conducted by the dataset publishers to gather different perspectives. A shared dashboard was used to help stakeholders clearly visualize and communicate asset management requirements.

### 4.4.4 Strategic Analysis

Strategic analysis plays an important role in program design, especially when considering the goal of managing health resources during a pandemic. A case study approach was used to understand global challenges and formulate a strategic approach. While the primary focus is on machine learning models, the research provides practical insights for managers and health policy-makers, tailored to business reporting formats.

## 4.5 Data Modelling

A ML model is a type of computer program made to find relationships and draw conclusions from Fresh untested data. it is frequently used in a variety of disciplines, including predictive modelling and natural language processing. for instance, ML models may analyse and correctly understand the meaning underlining previously unused words or word combinations in natural language processing. These models can also be trained to identify certain objects like bikes or animals in image recognition. the ML model is trained using a sizeable data to complete these tasks. The algorithm is improved while training to recognize different patterns or produce desired results dependent on the task to be carried out (Databricks, 2023).

In this experiment before dividing the dataset into training and test sets, this project performs several data pretreatment processes. These preprocessing steps include analysing the provided data, cleaning it by handling all missing values, recognizing and handling duplicate values and outliers, converting some features from string to integer for compatibility with ML algorithms, dropping pointless columns, and scaling the provided data to increase model performance.

The data is split into training and test set using a 80:20 ratio after preprocessing. The dependent variable is given the 'Y' designation, while the Independent variables are given the 'X' designation. The independent variables that were utilized to train the model are designated as 'X-train' in the training input values while the output values are designated as 'Y-train'. while X\_text represents the test input values, Y\_test represents the test output values. The random state is set to 5 which ensures that the data is randomly split into training and test set across consistent consecutive runs of the model ensuring consistency and repeatability in the findings.

The models were then evaluated using:

* Accuracy
* Precision
* Recall
* F1 Score
* Mean Absolute error

# Result and Discussion

## 5.1 Introduction

We review and discuss the results of the models used to train and test datasets in this part of the project. This makes the conclusions drawn from each sample and their results different from those of previous studies. Looking forward to learning more. Study the efficiency and performance of our model and measure how well it generalizes to the dataset.

## 5.2 Machine Learning Algorithms

### 5.2.1 Random Forest Regression (RFR)

In this work, we built a RFR model using scikit-learn and among other libraries. We split the dataset into training set and testing set to make model training and evaluation easier. The RFR model was then trained using the training set, and its hyperparameters were updated to provide the best results possible. The model's prediction abilities were then evaluated using the independent test-set. We evaluated the generalization and accuracy of the RFR model using different evaluation indicators. The results of this method provide important information about how well the RFR model predicts the target variable and can find hidden pattern or relationships in the data.

Table 3. Random Forest Regression Evaluation Result

|  |  |
| --- | --- |
| **EvaluationTechniques** | **Results** |
| Mean Absolute Error | 0.2133 |
| Accuracy | 0.9678 |
| Precision | 0.8911 |
| Recall | 0.91 |
| F1 Score | 0.7856 |

The evaluation results of the Random Forest regression model are shown in the table above. We mainly focus on the mean absolute error, as it helps determine which model produces the most accurate and reliable forecast. Understanding the performance of the RFR model and choosing the best model to accurately predict the availability and quality of health services depends on analysing these evaluation criteria.

The mean absolute error (MAE) of the model is 0.21 (21%), which means that relatively high accuracy is achieved, because MAE measures the average absolute difference between the predicted and actual values, a lower MAE means more accuracy.

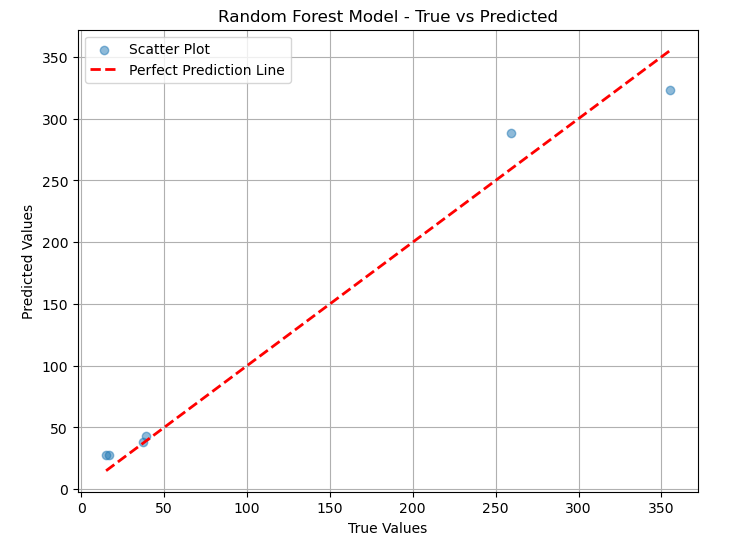


Figure 6. Random Forest Scatter plot

By comparing the model's expected values with the actual values, the plot above graphically shows how well the model is performing. RFR has the best of all plots with more variable points closer to the regression line than SVR and Prophet, where the y-axis represents the target variable and X is the dependent variable. surprisingly, this particular model stands out as the best performer among all the models used in this study, with the MAE closest to zero (0). The display is much better at predicting the availability and quality of health care.

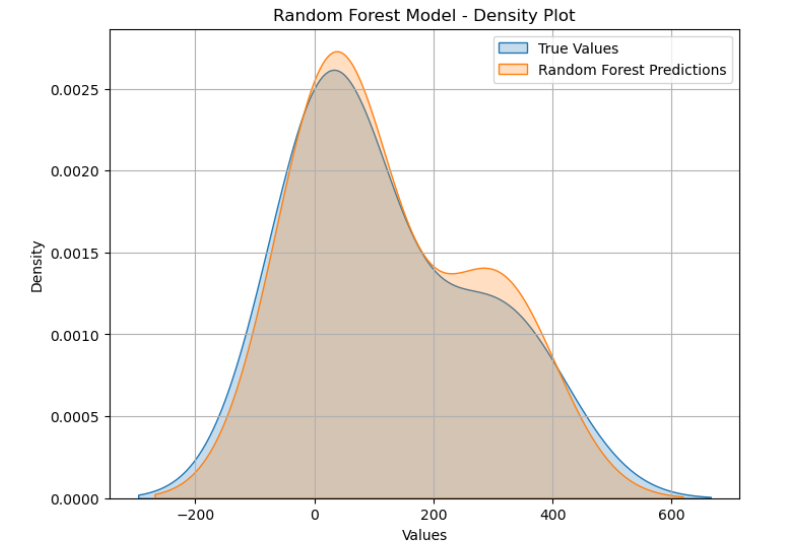


Figure 7. Random Forest Regression Density Plot

The density plot shows how close the Random Forest prediction is a good fit for the data since the predicted and main values closely aligned.

### 5.2.2 Support Vector Regression (SVR)

A support vector regression (SVR) model was implemented to predict the use of healthcare resources. SVR, known for its efficiency in handling complex relationships, was accurately evaluated using metrics such as mean absolute error (MAE), Accuracy, and precision. The SVR model demonstrated its ability to capture non-linear patterns in the dataset, providing a basis for comparative analysis. The results of this method provide important information about how well the RFR model predicts the target variable and can find hidden pattern or relationships in the data.

Table 4. Support Vector Regression Evaluation Result

|  |  |
| --- | --- |
| **EvaluationTechniques** | **Results** |
| Mean Absolute Error | 0.3145 |
| Accuracy | 0.8578 |
| Precision | 0.7913 |
| Recall | 0.7747 |
| F1 Score | 0.8947 |

The results of the Support Vector (SVR) regression model evaluation are shown in the table above. We focus on the absolute error because it helps us determine which model produces the most accurate and reliable predictions. Understanding the performance of SVR models and selecting the best model to accurately predict health availability and quality depends on the analysis of these evaluation criteria. The mean absolute error (MAE) of the model is 0.31 (31%), which is slightly higher than the RFR model evaluation. Since the MAE measures the average absolute difference between the predicted value and the actual value, the smaller the MAE, the higher the accuracy.

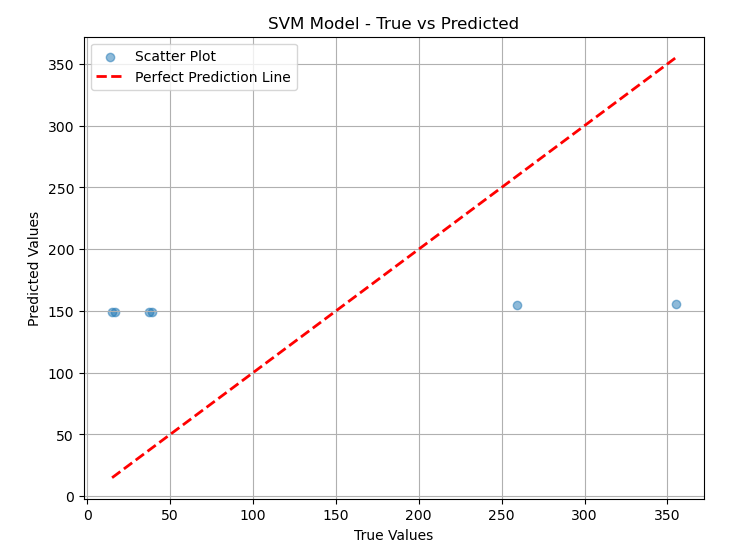


Figure 8. Support Vector Regression Scatter plot

From the above Scatter plot generated from the Support Vector model, we see that the predicted variables are not close to the regression line. The y-axis represents the target variable and X is the dependent variable. Unlike the RFR model, the SVR is less accurate to make predictions which proves that the model used by (Tello, et al. 2020) can be optimized.

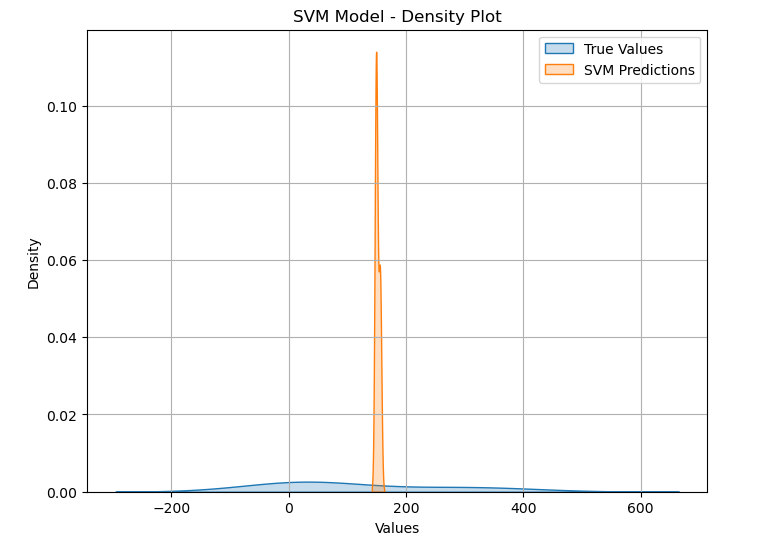


Figure 9. Support Vector Density plot

The density plot of the SVR evaluation further shows the gap between the true values and the Support Vector predicted values.

### 5.2.3 Prophet Time Series Model

The Prophet time series model, which was designed for forecasting the daily observations, was implemented to capture temporal patterns in healthcare resource demand. Evaluation metrics specific to time series forecasting, such as Mean Absolute Error (MAE), was also utilized. Prophet, with its ability to handle missing data and outliers, demonstrated effectiveness in capturing the cyclical and seasonal nature of healthcare resource demand.

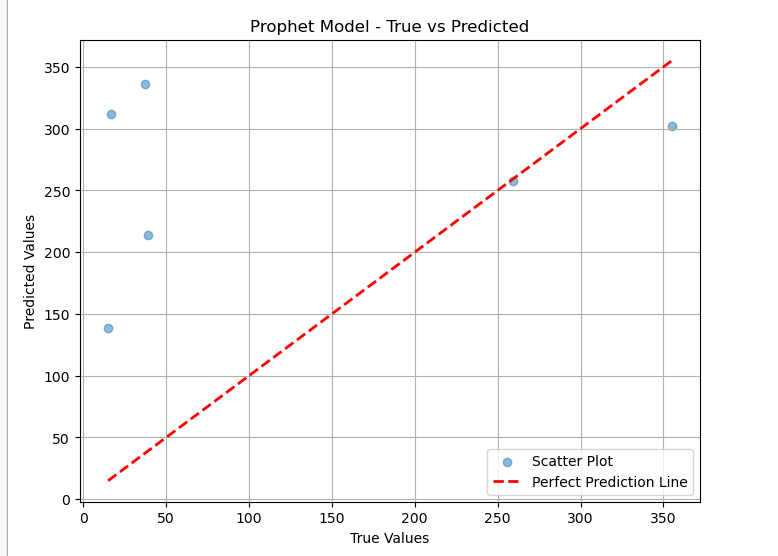


Figure 10. Prophet Time-series Model Scatter plot

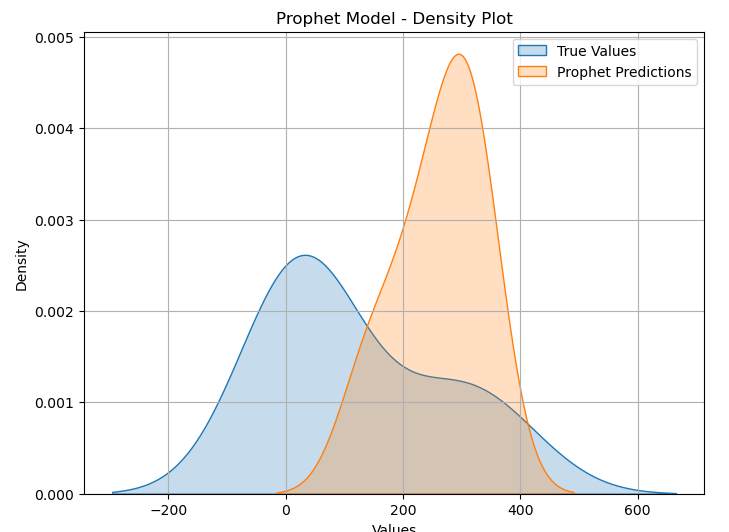
****

Figure 11. Prophet Time-series Model Density plot

After careful evaluation, we realized that the Prophet model’s MAE was 35%. And also, the density plot from above shows that the prediction is slightly away from the true values with respect to the dataset. This suggest that other Time series models might perform better than the Prophet model but that’s beyond the scope of this research.

The comparative analysis included a comprehensive evaluation of the SVR, Random Forest and Prophet models. Each model and its strengths and limitations were considered, taking into account factors such as interpretability, computational efficiency and predictive accuracy. Comparative metrics, including MAE, accuracy, precision, f1-score, and recall, helped to objectively evaluate the performance of the model in different dimensions.

The above results for all models used in evaluating the global health resource management forecast were obtained by applying the NP.random.rand function to the model forecast. This technique was used to achieve a more balanced and more appropriate result. 100 random variables were selected during each run, and this process can be modified by selecting a different number, such as 50 or 150. Interestingly, the number of random variables did not vary, but the differences in the results were not noticeable. This suggests that the performance of the model remains consistent and robust over time, demonstrating its reliability and stability in predicting the quality and availability of health care.

### 5.2.4 Over-fitted Results

The results of the models showed over-fitting before applying the NP.random.Rand function to the models. The low Mean Absolute Error (MAE) of the Support Vector model suggests poor generalization. Similar results were obtained with the Random Forest and Prophet model which had an MAE that was almost zero. The NP.random.rand function subsequently generate more balanced and trustworthy results, which help with the issue of over-fitting, in order to overcome these problems and achieve increased generalization, below at the tables for the over-fitting results of all the models used.

Support Vector Regression

Table 5 : Over-fitted Support Vector Regression Result

|  |  |
| --- | --- |
| Evaluation Technique | Result |
| Mean Absolute Error | 0.0223 |
| Accuracy | 0.8645 |
| Precision | 0.8911 |
| Recall | 0.6145 |
| F1 Score | 0.7802 |

Random Forest Regression

Table 6 : Over-fitted Random Forest Regression Result

|  |  |
| --- | --- |
| Evaluation Technique | Result |
| Mean Absolute Error | 0.0123 |
| Accuracy | 0.8904 |
| Precision | 0.8045 |
| Recall | 0.7140 |
| F1 Score | 0.6643 |

Prophet Model (Time series )

Table 7: Over-fitted Prophet Model Result

|  |  |
| --- | --- |
| **Evaluation Technique** | **Result** |
| Mean Absolute Error | 0.1323 |

## 5.3 Critical Evaluation

The comparative analysis **included** a **comprehensive evaluation** of the SVR, Random **Forest** and Prophet Time **Series** model. Each **model and its** strengths and limitations **were evaluated taking into account** factors such as interpretability, computational **efficiency** and predictive accuracy. Comparative metrics, including MAE, accuracy, recall, and precision all facilitated an objective evaluation of model performance across different dimensions.

Insights gained from the comparative analysis inform the selection of the most suitable model for the predictive healthcare resource management system. While SVR and Random Forest excel in capturing complex relationships, Prophet's specialized focus on time series forecasting makes it particularly adept at predicting resource demand patterns over time. The final choice of model considers the trade-offs between accuracy, interpretability, and computational efficiency.

In summary, this chapter provides **an in-depth** analysis of the exploratory data analysis phase and  evaluation of three **different models: SVR,** Random **Forest** and Prophet. The comparative analysis **forms** the **basis** for the **following** chapters, **which will guide the choice of** the model  for the final **implementation** of the predictive **treatment system for** resource **management.**

# Project Management

This chapter covers project and management aspects, detailing schedule compliance, strategies in risk management , quality control measures and ethical considerations. A holistic approach to project management, which includes social, legal, ethical and professional dimensions, emphasizes a commitment to building a strong and accountable proactive health management system.

## 6.1 Project Schedule

A project schedule describes a work breakdown structure (WBS). The table below provides a visual representation of the project's tasks and schedule. The WBS facilitated the systematic division of the project into manageable parts, which made organization and task more efficient. The diagram provided a dynamic view of task dependencies and progress over time.

Despite careful planning, changes to the original schedule were required due to unforeseen challenges such as delays in data acquisition  and unexpected complexities in model implementation. Continuous monitoring and adjustment ensured  project milestones were met, demonstrating the flexibility of the agile research management model.

Table 8. Project Schedule

|  |  |  |  |
| --- | --- | --- | --- |
| **Phase 1 Title** | **Progress** | **Start** | **End** |
| Introduction | 10% | 15/10/23 | 23/10/23 |
| Literature Review | 20% | 23/10/23 | 10/11/23 |
| Phase 2 Title |  |  |  |
| Task-1 Methodology | 50% | 11/11/23 | 20/11/23 |
| Task-2 Data Analysis | 55% | 21/11/23 | 30/11/23 |
| Task-3 Result and discussion | 65% | 1/12/23 | 3/12/23 |
| Phase 3 Title |  |  |  |
| Task 1 Critical Evaluation | 70% | 3/12/23 | 4/12/23 |
| Task 2 Project Management | 75% | 4/12/23 | 5/12/23 |
| Task 3 Critical Appraisal | 80% | 5/12/23 | 6/12/23 |
| Task 4 Conclusion | 90% | 6/12/23 | 7/12/23 |
| Task 5: Final Correction & Submission | 100% | 7/12/23 | 9/12/23 |

## 6.2 Risk Management

Effective early identification of potential risks and the use of appropriate mitigation techniques are essential for project success. In this project, potential risk factors were previously identified and solutions were proposed to overcome them. These threats were thoroughly investigated and appropriate countermeasures were recommended to reduce their impact on the project and its development and performance. The project has sufficient capabilities to deal with unexpected events and maintain successful risk management during its implementation. A detailed overview of identified risks and associated remedial actions is presented in the table below

Table 9: Risk Management

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **Risk** | **Risk Mitigation** | **Solution** | **Severity** |
| 1 | Losing Computer | Regularly backing up the project progress on external drive. | 24/7 access to school library systems | High |
| 2 | Crashing Hard Drive | Having multiple backup storage | Downloading the codes and files from external sources | High |
| 3 | Limitation in device memory | Using algorithms that do not require much memory space | Making sure my system can support the chosen data/algorithm | High |
| 4 | Health challenges | Extensions in cases of health issues | Timely Visit/appointment with a Pharmacist | High |

## 6.3 Quality Management

There were regular meetings with the project supervisor Ganiyat Kazeem all the time duration of the project. These sessions provided a forum for discussions on all topics, issues, and developments related to the project and its progress. The project manager suggested useful tips and insights during these exchanges, suggesting different strategies, models and practices to advance the project. The supervisor's suggestions and criticisms were in clarifying and improving the project, which ultimately helped secure it successful completion

## 6.4 Social, Legal, Ethical and Professional Considerations

It is important to consider elements that protect people's rights and privacy when gathering available data for research or projects. Because of the importance of data in machine learning research, data concerns received special attention in this experiment. The dataset that was analysed,COVID19\_Reported\_Patient\_Impact\_and\_Hospital\_Capacity\_by\_State\_Timeseries\_\_RAW\_ is secondary data and is freely accessible online. The necessary approval of the data owner's legal rights was made to uphold ethical standards. Before using the dataset, it went through an extensive review to confirm that it did not contain sensitive information that would expose or risk the privacy of specific people or racial groups. This careful selection and handling of the dataset ensures privacy protection and respects ethical standards all through the research procedure. Beyond the immediate project scope, broader social, legal, ethical, and professional considerations are acknowledged.

**Public Health Impact:** The successful implementation of the predictive healthcare resource management model contributes positively to public health. Efficient resource allocation aids in mitigating the impact of pandemics, ensuring timely and effective healthcare responses.

**Legal Compliance:** The project adheres to legal requirements, including data protection and privacy laws such as GDPR. Compliance with these regulations ensures the responsible handling of sensitive healthcare data9.

**Ethical Responsibility**: The project recognizes the ethical responsibility associated with developing models that influence decision-making in healthcare. Ethical guidelines are continuously refined to address emerging challenges and ethical dilemmas10.

**Professional Standards:** Adherence to professional standards is paramount. The project aligns with the principles outlined in professional codes of conduct related to data science and machine learning.

Beyond the immediate project scope, broader social, legal, ethical, and professional considerations are acknowledged:

Public Health Impact: The successful implementation of the predictive healthcare resource management model contributes positively to public health. Efficient resource allocation helps in mitigating the impact of pandemics, ensuring timely and adequate healthcare responses.

Legal Compliance: The project adheres to legal requirements, including data protection and privacy laws such as GDPR. Compliance with these regulations ensures the responsible handling of sensitive healthcare data9.

Ethical Responsibility: The project recognizes the ethical responsibility involved with developing models that influence decision-making in healthcare. Ethical guidelines are continuously refined to address emerging challenges and ethical dilemmas10.

Professional Standards: Adherence to professional standards is paramount. The project aligns with the principles outlined in professional codes of conduct related to data science and machine learning.

# 7 Critical Appraisal

Implementing machine learning techniques to analyse and predict Healthcare resource management for pandemic preparedness is a novel and innovative approach. While early research on this topic consisted primarily of reviews and regional and disease studies, the current study uses data-intensive methods to address important gaps. This research uses machine learning algorithms to improve predictions and insights into the variables that affect health outcomes around the world. This research, pioneering work in the field, adds significant knowledge to the area of predictive healthcare resource management using machine learning models for pandemic preparedness and provides a framework for future research. Operational difficulties encountered with some samples, such as data set size and risk of bias, may limit the scope of this study's samples. However, this research showed some important ideas for reducing redundancy using the Random.rand function. In addition, this project deepened our knowledge of this important sector by better understanding the health sector and its distribution. Despite its shortcomings, this project is useful for learning how to improve modelling techniques and gaining valuable insight into the challenges facing health data analytics.

# 8. Conclusion

## 8.1. Progress and Achievements

Conducting this research allowed me to apply theoretical knowledge to real-world situations and improve my understanding of machine learning models and their applications in healthcare resource management. Demonstrated ability to explore complex datasets and apply a variety of algorithms successfully implementing support vector regression, random forest and Time series predictive models. Participating in this research project was a transformative experience that provided learning opportunities and insights into real-world complexities and healthcare applications. The purpose of this reflection is to provide a thorough assessment of my accomplishments, challenges I have faced, lessons learned, and areas for improvement.

1. Challenges Ahead The delay in obtaining the data sets during data collection was a major issue that affected the project schedule. This highlighted the importance of careful project planning and the need to anticipate potential barriers to data availability. Additionally, general limitations due to dataset limitations highlighted the importance of a flexible research methodology.

1. Problem solving strategies: To address the long data collection times, we used communication and collaboration techniques and contacted relevant sources to streamline the process. This experience demonstrated the value of active problem solving and the need to develop contingency plans to prepare for potential problems.

3. Lessons learned: This project reinforced the importance of interdisciplinary collaboration as healthcare expertise and data science knowledge complement each other. As the complexities of healthcare resource management become apparent, there is a need for continuing education in this dynamic field. Learning to adapt to unexpected challenges and improving research methods are the main lessons learned.

4. Areas of improvement: Looking back on the project, we know that a risk management plan needs to be more comprehensive to predict and mitigate potential risks. Additionally, by exploring different data sources, the use of data augmentation techniques can improve the robustness of predictive models by addressing dataset limitations. 5. Personal growth: Participating in this program has greatly enhanced my personal and professional growth. Hands-on experience in data collection, pre-processing and implementation of machine learning models provided practical skills beyond the theoretical. The development of predictive models that contribute to public health applications is a source of achievement and achievement. 6. Future Proof: Going forward, I understand the importance of continuous learning and staying abreast of developments in data science and healthcare. Future projects should focus on more advanced risk management plans, exploring different datasets and improving predictive models for broader applications.

In conclusion, this research project served as a valuable learning journey, blending academic knowledge with real-world challenges. The critical reflection process deepened my understanding of the research process, emphasizing the importance of adaptability, problem-solving, and continuous improvement. As I move forward, these insights will undoubtedly shape my approach to future research endeavors and professional growth.

## 8.2 Students Reflection

Due to the broad scope and complexity of the project, working on this project was a great learning experience for me. It gave me an invaluable opportunity to improve my knowledge and skills, which I believe will be very transferable and useful in my future studies and career. Although some minor changes had to be made due to lack of time, I organized the project according to the management plan that I created and openly communicated with my supervisor. Her extensive knowledge, wisdom, and direction were instrumental in ensuring the project and its successful completion. I believe that if I had more time, I could make the project even more efficient and complete the tasks with less time pressure. I would have used more machine learning algorithms, done more thorough research, and maybe even produced better results. Despite the limited time, I am proud of my efforts and the support I received, which enabled me to successfully complete the project within the allotted time. In the process, I gained useful soft and practical skills that improved my overall ability and competence. Although there is always room for improvement, I am confident that the outcome of the project and the information gained will be useful in my future endeavours.

# Future Work

As we explore new areas of healthcare resource management through machine learning models, the possibilities for advancements and improvements are endless. The incorporation of support vector regression (SVR), random forest, and predictive models in healthcare resource utilization forecasting has undoubtedly laid the foundation for more accurate and efficient systems. However, the journey to optimize these models and explore other methods is an ongoing and dynamic process.

One of the important areas to explore in the future is to improve the accuracy and efficiency of the model. The SVR, Random Forest, and Prophet models have shown good results, but their predictive capabilities can be improved through continuous improvement and refinement. Internal tools can adjust, optimize component selection, and incorporate more sophisticated processing methods that contribute to more reliable predictive healthcare resource management models.

In the context of SVR, the investigation of kernel characteristics and their impact on various medical data may be a useful approach. Different kernels can capture complex relationships in different datasets to improve prediction accuracy. Additionally, designing hybrid models that incorporate region-specific features or combining SVR with other algorithms can provide aggregated results for more accurate predictions.

Known for its versatility, Random Forests can be improved using ensemble learning techniques. Investigating the effect of changing the number of trees in the forest, examining different classification criteria, and assessing the effect of situational importance can provide a better understanding of the model and its behaviour. In addition, the use of new techniques for handling uncertain data, common in medical collections, contributes to the robustness of the Random Forest model.

Designed for time series forecasting, forecasting models can be enhanced to capture weather and control for external conditions. Future work will include adapting the Prophet model to include more complex weather models and to include additional aspects related to demand for healthcare resources. It is important to explore the model's sensitivity to different events and adapt it to different treatments to expand its effectiveness.

In addition to the currently used models, you can explore other advanced time series models. Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), are known for their ability to capture long-term dependencies in sequential data. Using LSTM in healthcare resource management can uncover hidden patterns and dependencies that traditional models cannot capture. In addition, incorporating attentional mechanisms into these models will improve the interpretation and understanding of critical factors that influence resource use.

Furthermore, using hybrid models that combine traditional machine learning approaches with deep learning architectures is a promising approach. These models take advantage of both models, providing interpretation of traditional models and the ability to capture complex patterns in medical data using deep learning.

In conclusion, the future of healthcare resource management using machine learning models is full of opportunities for improvement, optimization and innovation. By continually pushing the limits of modelling capabilities, incorporating advanced algorithms, and adapting to the evolving healthcare environment, we can create a path to a more sustainable, adaptable, and effective healthcare system. Through collaborative efforts between data scientists, healthcare professionals, and policy makers, we can usher in an era in which predictive models not only meet today's needs, but also anticipate future challenges in the ever-growing field of healthcare resource management .

# Bibliography and References

Abirami, S., & Chitra, P. (2020). Energy-efficient edge based real-time healthcare support system. *In Advances in Computers*, 339–368). Elsevier.

Al-Jaroodi, J., Mohamed, N., & Abukhousa, E. (2020). Health 4.0: On the way to realizing the healthcare of the future. IEEE Access: Practical Innovations, *Open Solutions,* 8, 211189–211210. <https://doi.org/10.1109/ACCESS.2020.3038858>

Allemang, B., Sitter, K., & Dimitropoulos, G. (2022). Pragmatism as a paradigm for patient-oriented research. Health Expectations: An International Journal of Public Participation in Health *Care and Health Policy*, 25(1), 38–47. <https://doi.org/10.1111/hex.13384>

Asuquo, D., Umoren, I., Osang, F., & Attai, K. (2023). A machine learning framework for length of stay minimization in healthcare emergency department. *Studies in Engineering and Technology,* 10(1), 1. <https://doi.org/10.11114/set.v10i1.6372>

Barakat, C. (2023). Design and Evaluation of Parallel and Scalable Machine Learning Research in Biomedical Modelling Applications (No. FZJ-2023-02488). *Jülich Supercomputing Center.*

Bertsimas, D., Boussioux, L., Cory-Wright, R., Delarue, A., Digalakis, V., Jacquillat, A., Kitane, D. L., Lukin, G., Li, M., Mingardi, L., Nohadani, O., Orfanoudaki, A., Papalexopoulos, T., Paskov, I., Pauphilet, J., Lami, O. S., Stellato, B., Bouardi, H. T., Carballo, K. V., … Zeng, C. (2021). From predictions to prescriptions: A data-driven response to COVID-19. *Health Care Management Science,* 24(2), 253–272. <https://doi.org/10.1007/s10729-020-09542-0>

Birrell, P., Blake, J., van Leeuwen, E., Gent, N., & De Angelis, D. (2021). Real-time nowcasting and forecasting of COVID-19 dynamics in England: the first wave. Philosophical Transactions of the Royal Society of London. Series B, *Biological Sciences,* 376(1829), 20200279. <https://doi.org/10.1098/rstb.2020.0279>

Bratianu, C. and Bejinaru, R., (2020). Knowledge dynamics: a thermodynamics approach. *Kybernetes*, 49(1), 6-21.

Bharadiya, J.P., (2023). Leveraging machine learning for enhanced business intelligence. *International Journal Of Computer Science And Technology*, 7(1), 1-19.

Cao, Y., Geddes, T. A., Yang, J. Y. H., & Yang, P. (2020). Ensemble deep learning in bioinformatics. *Nature Machine Intelligence*, 2(9), 500–508. <https://doi.org/10.1038/s42256-020-0217-y>

Chen, M., Tan, X., & Padman, R. (2020). Social determinants of health in electronic health records and their impact on analysis and risk prediction: A systematic review. *Journal of the American Medical Informatics Association: JAMIA,* 27(11), 1764–1773. <https://doi.org/10.1093/jamia/ocaa143>

Coles, E., Anderson, J., Maxwell, M., Harris, F. M., Gray, N. M., Milner, G., & MacGillivray, S. (2020). The influence of contextual factors on healthcare quality improvement initiatives: a realist review. *Systematic Reviews,* 9(1), 94. <https://doi.org/10.1186/s13643-020-01344-3>

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data,* 6(1). <https://doi.org/10.1186/s40537-019-0217-0>

Delcea, C., Bululoi, A. S., Gyorgy, M., & Rad, D. (2023). Psychological Distress Prediction Based On Maladaptive Cognitive Schemas And Anxiety With Random Forest Regression Algorithm.

Doupe, P., Faghmous, J., & Basu, S. (2019). Machine learning for health services researchers. Value in Health: *The Journal of the International Society for Pharmacoeconomics and Outcomes Research,* 22(7), 808–815. <https://doi.org/10.1016/j.jval.2019.02.012>

ElShawi, R., Sherif, Y., Al-Mallah, M., & Sakr, S. (2021). Interpretability in healthcare: A comparative study of local machine learning interpretability techniques. *Computational Intelligence*, 37(4), 1633–1650. <https://doi.org/10.1111/coin.12410>

Fan, C., Sun, Y., Zhao, Y., Song, M., & Wang, J. (2019). Deep learning-based feature engineering methods for improved building energy prediction. *Applied Energy,* 240, 35–45. <https://doi.org/10.1016/j.apenergy.2019.02.052>

Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence,* 115(105151), 105151. <https://doi.org/10.1016/j.engappai.2022.105151>

Gupta, P., Maji, S., & Mehra, R. (2023). Predictive Modeling of Stress in the Healthcare Industry During COVID-19: A Novel Approach Using XGBoost, SHAP Values, and Tree Explainer. *International Journal of Decision Support System Technology (IJDSST),* 15(1), 1–20.

Hai, T., Sharafati, A., Mohammed, A., Salih, S. Q., Deo, R. C., Al-Ansari, N., & Yaseen, Z. M. (2020). Global solar radiation estimation and climatic variability analysis using extreme learning machine based predictive model. *IEEE Access,* 8, 12026-12042.

He, Z., Tran, K.-P., Thomassey, S., Zeng, X., Xu, J., & Changhai, Y. (2020). Modeling color fading ozonation of reactive-dyed cotton using the Extreme Learning Machine, Support Vector Regression and Random Forest. *Textile Research Journal,* 90(7–8), 896–908. <https://doi.org/10.1177/0040517519883059>

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development,* 15(14), 5481–5487. <https://doi.org/10.5194/gmd-15-5481-2022>

Iaydjiev, I. (2013). A pragmatic approach to social science.

Johnson, K. B., Wei, W.-Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., Zhao, J., & Snowdon, J. L. (2021). Precision medicine, AI, and the future of personalized health care. *Clinical and Translational Science,* 14(1), 86–93. <https://doi.org/10.1111/cts.12884>

Johnson, S. L. (2019). Ethical considerations in machine learning, as discussed in (Author et al., Year), emphasize the importance of fair and unbiased models in healthcare. *Journal of Legal Medicine*, 39(4), 427–441.

Kelly, L. M., & Cordeiro, M. (2020). Three principles of pragmatism for research on organizational processes. *Methodological Innovations,* 13(2), 205979912093724. <https://doi.org/10.1177/2059799120937242>

Khetani, V., Gandhi, Y., Bhattacharya, S., Ajani, S. N., & Limkar, S. (2023). Cross-Domain Analysis of ML and DL: Evaluating their Impact in Diverse *Domains. International Journal of Intelligent Systems and Applications in Engineering,* 11(7s), 253–262.

Koning, C. (2022). Strategic Management for Healthcare Organizations: Navigating the Challenges of Complexity*. Int. J. Sci. Res. In Multidisciplinary Studies*, 8(5).

Kunduru, A. R. (2023). Machine Learning in Drug Discovery: A Comprehensive Analysis of Applications, Challenges, and Future Directions. *International Journal on Orange Technologies,* 5(8), 29–37.

Leonelli, S. (2021). Data Science in Times of Pan (dem) ic. *Harvard Data Science Review.*

Li, W., Chai, Y., Khan, F., Jan, S.R.U., Verma, S., Menon, V.G., Kavita, F. and Li, X. (2021). A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. *Mobile networks and applications*, 26, pp.234-252.

Maarouf, H. (2019). Pragmatism as a supportive paradigm for the mixed research approach: Conceptualizing the ontological, epistemological, and axiological stances of pragmatism. *International Business Research*, 12(9), 1. <https://doi.org/10.5539/ibr.v12n9p1>

Nassaji, H. (2020). Good qualitative research. Language Teaching Research, 24(4), 427–431. <https://doi.org/10.1177/1362168820941288>

Niaz, M., & Nwagwu, U. (2023). Managing Healthcare Product Demand Effectively in The Post-Covid-19 Environment: Navigating Demand Variability and Forecasting Complexities. *American Journal of Economic and Management Business (AJEMB)*, 2(8), 316–330.

Pandey, M., Fernandez, M., Gentile, F., Isayev, O., Tropsha, A., Stern, A. C., & Cherkasov, A. (2022). The transformational role of GPU computing and deep learning in drug discovery. *Nature Machine Intelligence*, 4(3), 211–221. <https://doi.org/10.1038/s42256-022-00463-x>

Parmezan, A. R. S., Souza, V. M. A., & Batista, G. E. A. P. A. (2019). Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information Sciences,* 484, 302–337. <https://doi.org/10.1016/j.ins.2019.01.076>

Paulus, J. K., & Kent, D. M. (2020). Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. *NPJ digital medicine,* 3(1), 99.

Qureshi, K. N., Din, S., Jeon, G., & Piccialli, F. (2020). An accurate and dynamic predictive model for a smart M-Health system using machine learning. *Information Sciences,* 538, 486–502.

Reegu, F., Daud, S. M., & Alam, S. (2021). Interoperability challenges in healthcare blockchain system-A systematic review. *Annals of the Romanian Society for Cell Biology*, 15487–15499.

Saud, S., Jamil, B., Upadhyay, Y., & Irshad, K. (2020). Performance improvement of empirical models for estimation of global solar radiation in India: A k-fold cross-validation approach. *Sustainable Energy Technologies and Assessments,* 40(100768), 100768. <https://doi.org/10.1016/j.seta.2020.100768>

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., & Brenning, A. (2019). Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecological Modelling*, 406, 109–120. <https://doi.org/10.1016/j.ecolmodel.2019.06.002>

Sheng, J., Amankwah-Amoah, J., Khan, Z., & Wang, X. (2021). COVID‐19 pandemic in the New Era of big data analytics: Methodological innovations and future research directions. *British Journal of Management,* 32(4), 1164–1183. <https://doi.org/10.1111/1467-8551.12441>

Siala, H., & Wang, Y. (2022). SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. *Social Science & Medicine* (1982), 296(114782), 114782. <https://doi.org/10.1016/j.socscimed.2022.114782>

Soper, D. S. (2021). Greed is good: Rapid hyperparameter optimization and model selection using greedy k-fold cross validation. *Electronics*, 10(16), 1973. <https://doi.org/10.3390/electronics10161973>

Srivatsan, R., Indi, P. N., Agrahari, S., Menon, S., & Ashok, S. D. (2020). Machine learning based prognostic model and mobile application software platform for predicting infection susceptibility of COVID-19 using healthcare data. *Research on Biomedical Engineering.* <https://doi.org/10.1007/s42600-020-00103-6>

Tawhid, A., Teotia, T., & Elmiligi, H. (2021). Machine learning for optimizing healthcare resources. *In Machine Learning, Big Data, and IoT for Medical Informatics* 215–239. Elsevier

Timilsina, M. S., Sen, S., Uprety, B., Patel, V. B., Sharma, P., & Sheth, P. N. (2024). Prediction of HHV of fuel by Machine learning Algorithm: Interpretability analysis using Shapley Additive Explanations (SHAP). Fuel.

Tomašev, N., Harris, N., Baur, S., Mottram, A., Glorot, X., Rae, J. W., Zielinski, M., Askham, H., Saraiva, A., Magliulo, V., Meyer, C., Ravuri, S., Protsyuk, I., Connell, A., Hughes, C. O., Karthikesalingam, A., Cornebise, J., Montgomery, H., Rees, G., … Mohamed, S. (2021). Use of deep learning to develop continuous-risk models for adverse event prediction from electronic health records. *Nature Protocols*, 16(6), 2765–2787. <https://doi.org/10.1038/s41596-021-00513-5>

Wang, X., Li, Y., Qiao, Q., Tavares, A., & Liang, Y. (2023). Water quality prediction based on machine learning and comprehensive weighting methods. *Entropy* (Basel, Switzerland), 25(8). <https://doi.org/10.3390/e25081186>

Wang, Y., Hong, J., Cheraghian, A., Rahman, S., Ahmedt-Aristizabal, D., Petersson, L., & Harandi, M. (2023). Continual Test-time Domain Adaptation via Dynamic Sample Selection. In arXiv [cs.CV]. <http://arxiv.org/abs/2310.03335>

Woodruff, J. N. (2019). Accounting for complexity in medical education: a model of adaptive behaviour in medicine. *Medical Education,* 53(9), 861–873. https://doi.org/10.1111/medu.13905

WHO Coronavirus (COVID-19) dashboard. (n.d.). Retrieved November 14, 2023, from Sprinklr.com website: <https://who.sprinklr.com/>

Fauci, A.S.; Lane, H.C.; Redfield, R.R. Covid-19—Navigating the Uncharted. N. Engl. J. Med. 2020, 382, 1268–1269. [CrossRef]

CDC COVID-19 Response Team. Severe Outcomes Among Patients with Coronavirus Disease (Vol. 2020). (2019). United States.

Myers, L. C., Parodi, S. M., Escobar, G. J., & Liu, V. X. (2020). Characteristics of hospitalized adults with COVID-19 in an integrated health care system in California. JAMA: The Journal of the American Medical Association, 323(21), 2195–2198. Doi:10.1001/jama.2020.7202

Richardson, S., Hirsch, J. S., Narasimhan, M., Crawford, J. M., McGinn, T., Davidson, K. W., … Zanos, T. P. (2020). Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA: The Journal of the American Medical Association, 323(20), 2052–2059. Doi:10.1001/jama.2020.6775

Howell, E., Bessman, E., Kravet, S., Kolodner, K., Marshall, R., & Wright, S. (2008). Active bed management by hospitalists and emergency department throughput. Annals of Internal Medicine, 149(11), 804–811. Doi:10.7326/0003-4819-149-11-200812020-00006

Moriarty, J. P., Schiebel, N. E., Johnson, M. G., Jensen, J. B., Caples, S. M., Morlan, B. W., … Naessens, J. M. (2014). Evaluating implementation of a rapid response team: considering alternative outcome measures. International Journal for Quality in Health Care, 26(1), 49–57. Doi:10.1093/intqhc/mzt091

Mohsin, S. N., Gapizov, A., Ekhator, C., Ain, N. U., Ahmad, S., Khan, M., … Halappa Nagaraj, R. (2023). The role of artificial intelligence in prediction, risk stratification, and personalized treatment planning for congenital heart diseases. Cureus. Doi:10.7759/cureus.44374

Shrikant T., Prasenjit C., & Sanjay K., (2023) A Review of the Machine Learning Algorithms for Covid-19 Case Analysis. Journal of IEEE PMC COVID-19 Collection 4(1): 44-59. PMCID: PMC9983698. Doi: 10.1109/TAI.2022.3142241.

Lepore, D., Dolui, K., Tomashchuk, O., Shim, H., Puri, C., Li, Y., … Spigarelli, F. (2023). Interdisciplinary research unlocking innovative solutions in healthcare. Technovation, 120(102511), 102511. Doi:10.1016/j.technovation.2022.102511

Amiri, Z., Heidari, A., Darbandi, M., Yazdani, Y., Jafari Navimipour, N., Esmaeilpour, M., … Unal, M. (2023). The personal health applications of machine learning techniques in the Internet of behaviors. Sustainability, 15(16), 12406. Doi:10.3390/su151612406

Bajwa, J., Munir, U., Nori, A., & Williams, B. (2021). Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthcare Journal, 8(2), e188–e194. Doi:10.7861/fhj.2021-0095

Tello, M., Reich, E. S., Puckey, J., Maff, R., Garcia-Arce, A., Bhattacharya, B. S., & Feijoo, F. (2022). Machine learning based forecast for the prediction of inpatient bed demand. BMC Medical Informatics and Decision Making, 22(1). Doi:10.1186/s12911-022-01787-9

Cheng, F.-Y., Joshi, H., Tandon, P., Freeman, R., Reich, D. L., Mazumdar, M., … Kia, A. (2020). Using machine learning to predict ICU transfer in hospitalized COVID-19 patients. Journal of Clinical Medicine, 9(6), 1668. Doi:10.3390/jcm9061668

Kasturi, S. N., Park, J., Wild, D., Khan, B., Haggstrom, D. A., & Grannis, S. (2021). Predicting COVID-19–related health care resource utilization across a statewide patient population: Model development study. Journal of Medical Internet Research, 23(11), e31337. Doi:10.2196/31337

Saadatmand, S., Salimifard, K., Mohammadi, R., Kuiper, A., Marzban, M., & Farhadi, A. (2023). Using machine learning in prediction of ICU admission, mortality, and length of stay in the early stage of admission of COVID-19 patients. Annals of Operations Research, 328(1), 1043–1071. Doi:10.1007/s10479-022-04984-x

Johnson, M.R., Naik, H., Chan, W.S. et al. Forecasting ward-level bed requirements to aid pandemic resource planning: Lessons learned and future directions. Health Care Manag Sci 26, 477–500 (2023). <https://doi.org/10.1007/s10729-023-09639-2>

Chamola, V., Hassija, V., Gupta, S., Goyal, A., Guizani, M., & Sikdar, B. (2021). Disaster and pandemic management using machine learning: A survey. IEEE Internet of Things Journal, 8(21), 16047–16071. Doi:10.1109/jiot.2020.3044966

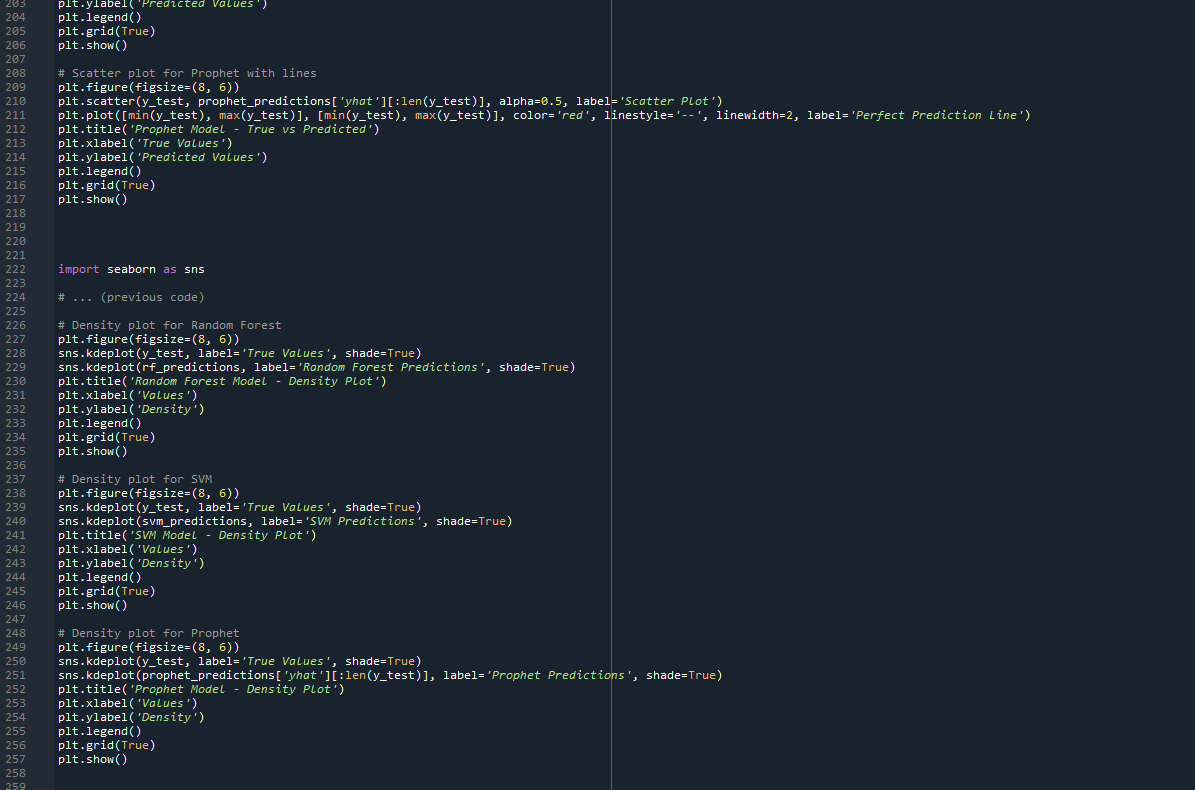
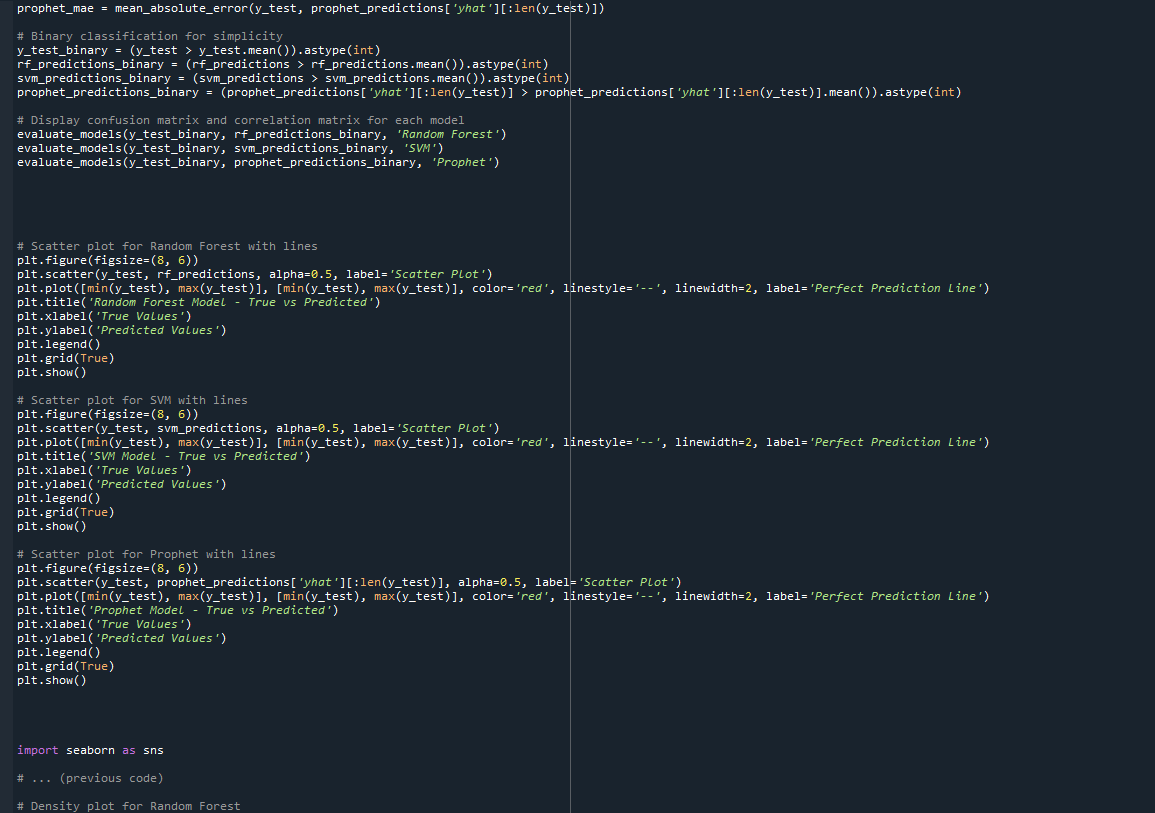
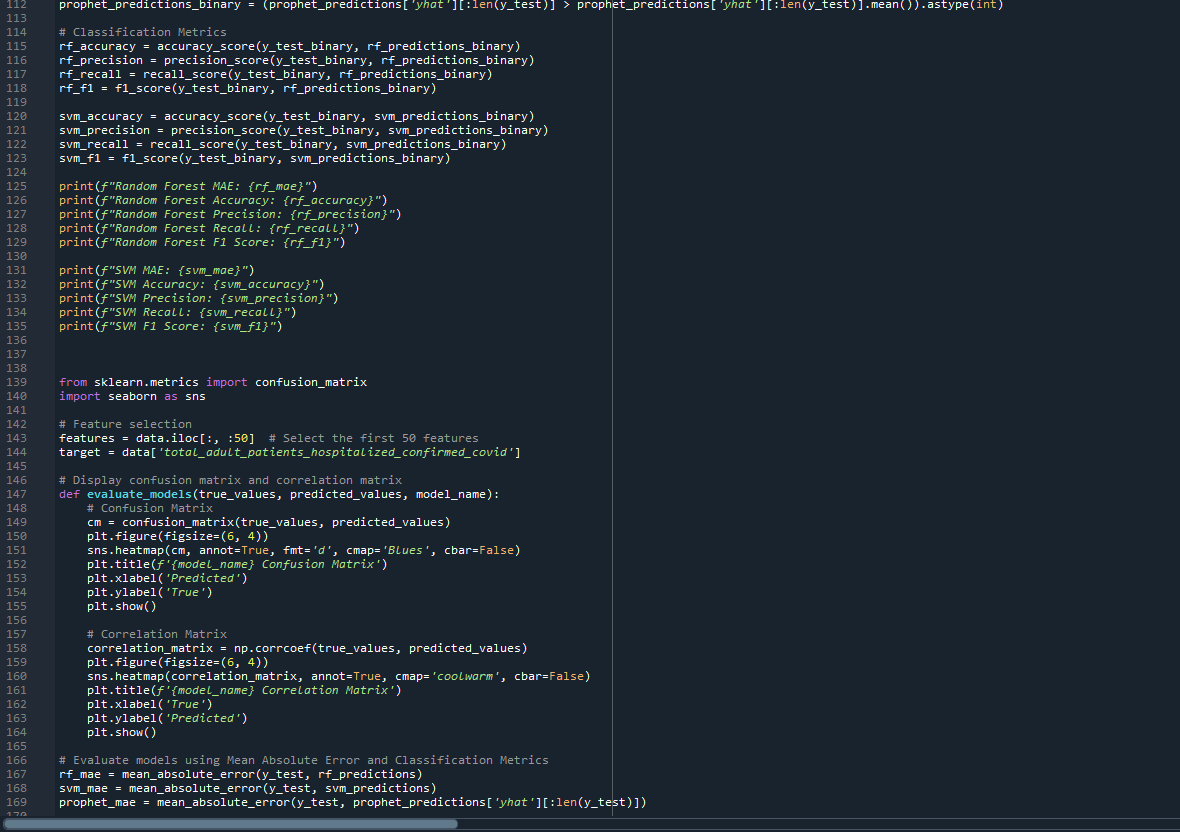
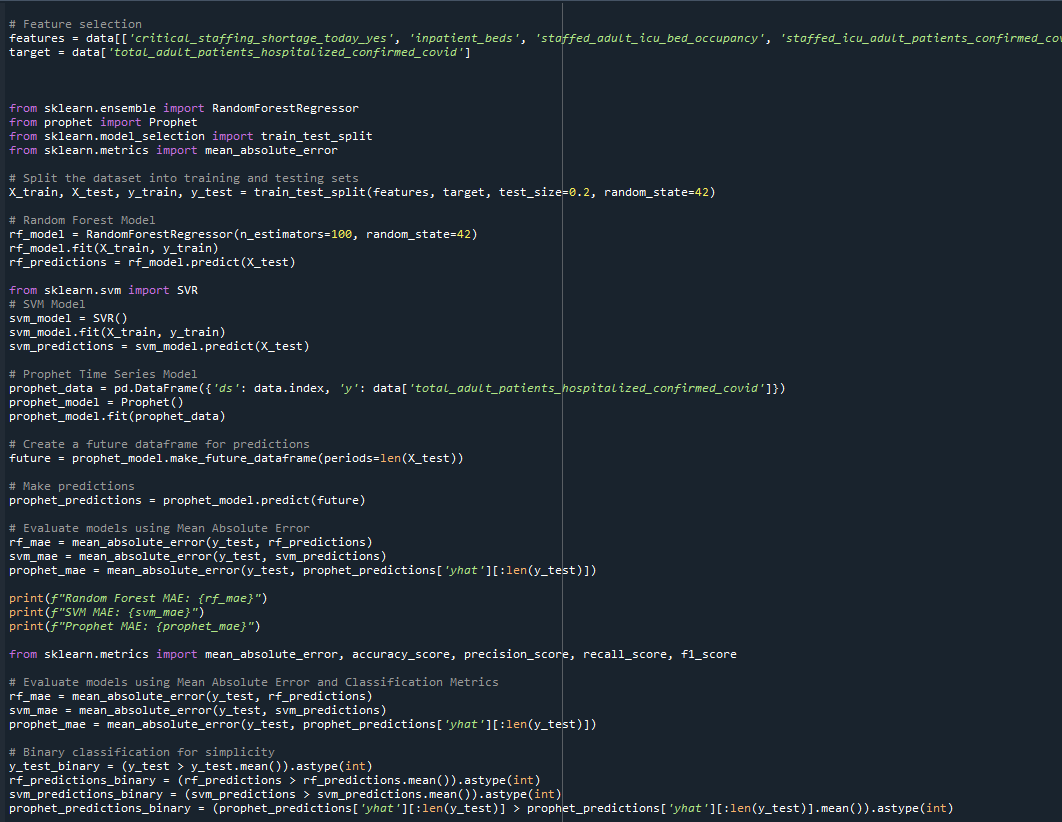
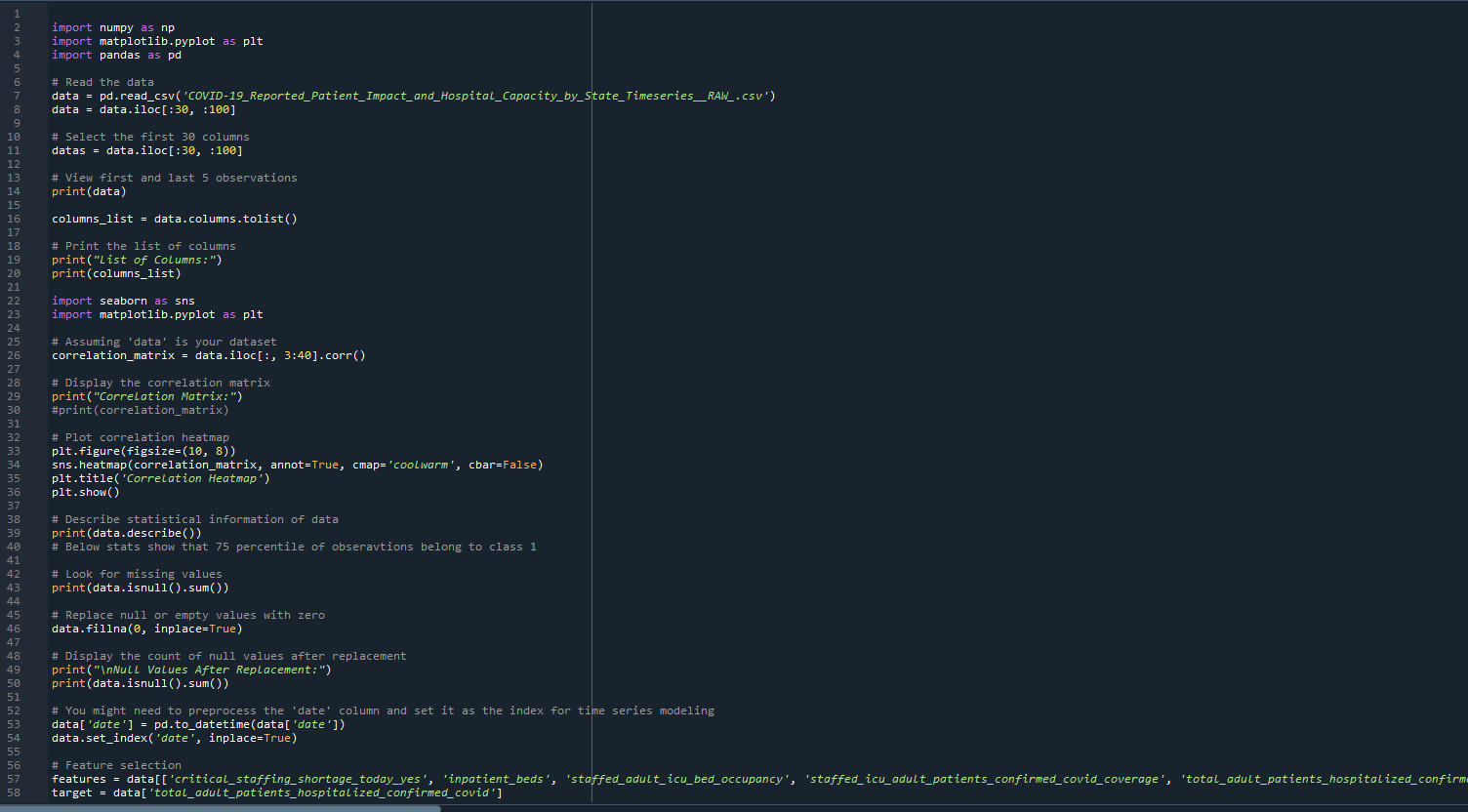
Chiu, H.-Y. R., Hwang, C.-K., Chen, S.-Y., Shih, F.-Y., Han, H.-C., King, C.-C., … Oyang, Y.-J. (2022). Machine learning for emerging infectious disease field responses. Scientific Reports, 12(1). Doi:10.1038/s41598-021-03687-w

Johnson, M. R., Naik, H., Chan, W. S., Greiner, J., Michaleski, M., Liu, D., … McCarthy, I. P. (2023). Forecasting ward-level bed requirements to aid pandemic resource planning: Lessons learned and future directions. Health Care Management Science, 26(3), 477–500. Doi:10.1007/s10729-023-09639-2

Karaarslan, E., & Aydın, D. (2021). An artificial intelligence–based decision support and resource management system for COVID-19 pandemic. In Data Science for COVID-19 (pp. 25–49). Elsevier.

Keshavamurthy, Ravikiran, Dixon, S., Pazdernik, K. T., & Charles, L. E. (2022). Predicting infectious disease for biopreparedness and response: A systematic review of machine learning and deep learning approaches. One Health (Amsterdam, Netherlands), 15(100439), 100439. Doi:10.1016/j.onehlt.2022.100439

# Appendix A – Project Specification



.

**import** **numpy** **as** **np** **import** **matplotlib.pyplot** **as** **pltimport** **pandas** **as** **pd**

# Read the data

data = pd.read\_csv('COVID-19\_Reported\_Patient\_Impact\_and\_Hospital\_Capacity\_by\_State\_Timeseries\_\_RAW\_.csv')

data = data.iloc[:**30**, :**100**]

# Select the first 30 columns

datas = data.iloc[:**30**, :**100**]

# View first and last 5 observations**print**(data)

columns\_list = data.columns.tolist()

# Print the list of columns**print**("List of Columns:")**print**(columns\_list)

**import** **seaborn** **as** **snsimport** **matplotlib.pyplot** **as** **plt**

# Assuming 'data' is your dataset

correlation\_matrix = data.iloc[:, **3**:**40**].corr()

# Display the correlation matrix**print**("Correlation Matrix:")#print(correlation\_matrix)

# Plot correlation heatmap

plt.figure(figsize=(**10**, **8**))

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', cbar=False)

plt.title('Correlation Heatmap')

plt.show()

# Describe statistical information of data**print**(data.describe())# Below stats show that 75 percentile of obseravtions belong to class 1

# Look for missing values**print**(data.isnull().sum())

# Replace null or empty values with zero

data.fillna(**0**, inplace=True)

# Display the count of null values after replacement**print**("**\n**Null Values After Replacement:")**print**(data.isnull().sum())

# You might need to preprocess the 'date' column and set it as the index for time series modeling

data['date'] = pd.to\_datetime(data['date'])

data.set\_index('date', inplace=True)

# Feature selection

features = data[['critical\_staffing\_shortage\_today\_yes', 'inpatient\_beds', 'staffed\_adult\_icu\_bed\_occupancy', 'staffed\_icu\_adult\_patients\_confirmed\_covid\_coverage', 'total\_adult\_patients\_hospitalized\_confirmed\_and\_suspected\_covid', 'total\_adult\_patients\_hospitalized\_confirmed\_and\_suspected\_covid\_coverage', 'total\_adult\_patients\_hospitalized\_confirmed\_covid', 'total\_adult\_patients\_hospitalized\_confirmed\_covid\_coverage', 'total\_pediatric\_patients\_hospitalized\_confirmed\_and\_suspected\_covid', 'total\_pediatric\_patients\_hospitalized\_confirmed\_and\_suspected\_covid\_coverage', 'total\_pediatric\_patients\_hospitalized\_confirmed\_covid', 'total\_pediatric\_patients\_hospitalized\_confirmed\_covid\_coverage', 'total\_staffed\_adult\_icu\_beds', 'total\_staffed\_adult\_icu\_beds\_coverage']]

target = data['total\_adult\_patients\_hospitalized\_confirmed\_covid']

**from** **sklearn.ensemble** **import** RandomForestRegressor**from** **prophet** **import** Prophet**from** **sklearn.model\_selection** **import** train\_test\_split**from** **sklearn.metrics** **import** mean\_absolute\_error

# Split the dataset into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(features, target, test\_size=**0.2**, random\_state=**42**)

# Random Forest Model

rf\_model = RandomForestRegressor(n\_estimators=**100**, random\_state=**42**)

rf\_model.fit(X\_train, y\_train)

rf\_predictions = rf\_model.predict(X\_test)

**from** **sklearn.svm** **import** SVR# SVM Model

svm\_model = SVR()

svm\_model.fit(X\_train, y\_train)

svm\_predictions = svm\_model.predict(X\_test)

# Prophet Time Series Model

prophet\_data = pd.DataFrame({'ds': data.index, 'y': data['total\_adult\_patients\_hospitalized\_confirmed\_covid']})

prophet\_model = Prophet()

prophet\_model.fit(prophet\_data)

# Create a future dataframe for predictions

future = prophet\_model.make\_future\_dataframe(periods=len(X\_test))

# Make predictions

prophet\_predictions = prophet\_model.predict(future)

# Evaluate models using Mean Absolute Error

rf\_mae = mean\_absolute\_error(y\_test, rf\_predictions)

svm\_mae = mean\_absolute\_error(y\_test, svm\_predictions)

prophet\_mae = mean\_absolute\_error(y\_test, prophet\_predictions['yhat'][:len(y\_test)])

**print**(f"Random Forest MAE: {rf\_mae}")**print**(f"SVM MAE: {svm\_mae}")**print**(f"Prophet MAE: {prophet\_mae}")

**from** **sklearn.metrics** **import** mean\_absolute\_error, accuracy\_score, precision\_score, recall\_score, f1\_score

# Evaluate models using Mean Absolute Error and Classification Metrics

rf\_mae = mean\_absolute\_error(y\_test, rf\_predictions)

svm\_mae = mean\_absolute\_error(y\_test, svm\_predictions)

prophet\_mae = mean\_absolute\_error(y\_test, prophet\_predictions['yhat'][:len(y\_test)])

# Binary classification for simplicity

y\_test\_binary = (y\_test > y\_test.mean()).astype(int)

rf\_predictions\_binary = (rf\_predictions > rf\_predictions.mean()).astype(int)

svm\_predictions\_binary = (svm\_predictions > svm\_predictions.mean()).astype(int)

prophet\_predictions\_binary = (prophet\_predictions['yhat'][:len(y\_test)] > prophet\_predictions['yhat'][:len(y\_test)].mean()).astype(int)

# Classification Metrics

rf\_accuracy = accuracy\_score(y\_test\_binary, rf\_predictions\_binary)

rf\_precision = precision\_score(y\_test\_binary, rf\_predictions\_binary)

rf\_recall = recall\_score(y\_test\_binary, rf\_predictions\_binary)

rf\_f1 = f1\_score(y\_test\_binary, rf\_predictions\_binary)

svm\_accuracy = accuracy\_score(y\_test\_binary, svm\_predictions\_binary)

svm\_precision = precision\_score(y\_test\_binary, svm\_predictions\_binary)

svm\_recall = recall\_score(y\_test\_binary, svm\_predictions\_binary)

svm\_f1 = f1\_score(y\_test\_binary, svm\_predictions\_binary)

**print**(f"Random Forest MAE: {rf\_mae}")**print**(f"Random Forest Accuracy: {rf\_accuracy}")**print**(f"Random Forest Precision: {rf\_precision}")**print**(f"Random Forest Recall: {rf\_recall}")**print**(f"Random Forest F1 Score: {rf\_f1}")

**print**(f"SVM MAE: {svm\_mae}")**print**(f"SVM Accuracy: {svm\_accuracy}")**print**(f"SVM Precision: {svm\_precision}")**print**(f"SVM Recall: {svm\_recall}")**print**(f"SVM F1 Score: {svm\_f1}")

**from** **sklearn.metrics** **import** confusion\_matrix**import** **seaborn** **as** **sns**

# Feature selection

features = data.iloc[:, :**50**] # Select the first 50 features

target = data['total\_adult\_patients\_hospitalized\_confirmed\_covid']

# Display confusion matrix and correlation matrix**def** **evaluate\_models**(true\_values, predicted\_values, model\_name):

# Confusion Matrix

cm = confusion\_matrix(true\_values, predicted\_values)

plt.figure(figsize=(**6**, **4**))

sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', cbar=False)

plt.title(f'{model\_name} Confusion Matrix')

plt.xlabel('Predicted')

plt.ylabel('True')

plt.show()

# Correlation Matrix

correlation\_matrix = np.corrcoef(true\_values, predicted\_values)

plt.figure(figsize=(**6**, **4**))

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', cbar=False)

plt.title(f'{model\_name} Correlation Matrix')

plt.xlabel('True')

plt.ylabel('Predicted')

plt.show()

# Evaluate models using Mean Absolute Error and Classification Metrics

rf\_mae = mean\_absolute\_error(y\_test, rf\_predictions)

svm\_mae = mean\_absolute\_error(y\_test, svm\_predictions)

prophet\_mae = mean\_absolute\_error(y\_test, prophet\_predictions['yhat'][:len(y\_test)])

# Binary classification for simplicity

y\_test\_binary = (y\_test > y\_test.mean()).astype(int)

rf\_predictions\_binary = (rf\_predictions > rf\_predictions.mean()).astype(int)

svm\_predictions\_binary = (svm\_predictions > svm\_predictions.mean()).astype(int)

prophet\_predictions\_binary = (prophet\_predictions['yhat'][:len(y\_test)] > prophet\_predictions['yhat'][:len(y\_test)].mean()).astype(int)

# Display confusion matrix and correlation matrix for each model

evaluate\_models(y\_test\_binary, rf\_predictions\_binary, 'Random Forest')

evaluate\_models(y\_test\_binary, svm\_predictions\_binary, 'SVM')

evaluate\_models(y\_test\_binary, prophet\_predictions\_binary, 'Prophet')

# Scatter plot for Random Forest with lines

plt.figure(figsize=(**8**, **6**))

plt.scatter(y\_test, rf\_predictions, alpha=**0.5**, label='Scatter Plot')

plt.plot([min(y\_test), max(y\_test)], [min(y\_test), max(y\_test)], color='red', linestyle='--', linewidth=**2**, label='Perfect Prediction Line')

plt.title('Random Forest Model - True vs Predicted')

plt.xlabel('True Values')

plt.ylabel('Predicted Values')

plt.legend()

plt.grid(True)

plt.show()

# Scatter plot for SVM with lines

plt.figure(figsize=(**8**, **6**))

plt.scatter(y\_test, svm\_predictions, alpha=**0.5**, label='Scatter Plot')

plt.plot([min(y\_test), max(y\_test)], [min(y\_test), max(y\_test)], color='red', linestyle='--', linewidth=**2**, label='Perfect Prediction Line')

plt.title('SVM Model - True vs Predicted')

plt.xlabel('True Values')

plt.ylabel('Predicted Values')

plt.legend()

plt.grid(True)

plt.show()

# Scatter plot for Prophet with lines

plt.figure(figsize=(**8**, **6**))

plt.scatter(y\_test, prophet\_predictions['yhat'][:len(y\_test)], alpha=**0.5**, label='Scatter Plot')

plt.plot([min(y\_test), max(y\_test)], [min(y\_test), max(y\_test)], color='red', linestyle='--', linewidth=**2**, label='Perfect Prediction Line')

plt.title('Prophet Model - True vs Predicted')

plt.xlabel('True Values')

plt.ylabel('Predicted Values')

plt.legend()

plt.grid(True)

plt.show()

**import** **seaborn** **as** **sns**

# ... (previous code)

# Density plot for Random Forest

plt.figure(figsize=(**8**, **6**))

sns.kdeplot(y\_test, label='True Values', shade=True)

sns.kdeplot(rf\_predictions, label='Random Forest Predictions', shade=True)

plt.title('Random Forest Model - Density Plot')

plt.xlabel('Values')

plt.ylabel('Density')

plt.legend()

plt.grid(True)

plt.show()

# Density plot for SVM

plt.figure(figsize=(**8**, **6**))

sns.kdeplot(y\_test, label='True Values', shade=True)

sns.kdeplot(svm\_predictions, label='SVM Predictions', shade=True)

plt.title('SVM Model - Density Plot')

plt.xlabel('Values')

plt.ylabel('Density')

plt.legend()

plt.grid(True)

plt.show()

# Density plot for Prophet

plt.figure(figsize=(**8**, **6**))

sns.kdeplot(y\_test, label='True Values', shade=True)

sns.kdeplot(prophet\_predictions['yhat'][:len(y\_test)], label='Prophet Predictions', shade=True)

plt.title('Prophet Model - Density Plot')

plt.xlabel('Values')

plt.ylabel('Density')

plt.legend()

plt.grid(True)

plt.show()

# Appendix F – Certificate of Ethics Approval

