Arab Open University- Egypt



Faculty of Computer Studies

Information Technology and Computing Department

**Predictive Analytics for Hospital**

**Bed Occupancy Management**

Ali abdelmeneam - 21510563

TM471: Final Year Project, tbd  
Supervisor: Dr Amira fawzy

**Abstract**

The "Predictive Analytics for Hospital Bed Occupancy Management" project tries to build a machine learning based predictive model to predict hospital bed occupancy using historical patient admission historical data.  
  
The solution's aim is to meet the needs of hospitals for optimization of resources (e.g., overcrowding of inpatient units and under-use of beds). By leveraging data analytics and a user-friendly web-based dashboard.  
  
The project offers hospital administrators with clinically relevant actionable information for informed resource decisions. This framework is based on using Python libraries for data preprocessing, modeling, and visualization, thus providing an accurate and applicable view of the management of hospital resources and enhancement of patient care results.

**Acknowledgment**

I would like to express my sincere gratitude to my supervisor, Dr. Amira fawzy for providing me with her invaluable guidance, comments and suggestions throughout the course of this project. I would also like to thank her for motivating me to work harder to make this project successful. Her suggestions and time was most valuable and appreciated.

Secondly, I would like to thank the astoundingly supportive staff at the AOU for the resources and top-notch facilities given to me to achieve and advance my learning in Informational Technology and Computing.

Moreover, I’m proud to be part of a highly intelligent team of students that partnered with me to research and construct this project. The integrity of my teammates was clearly shown by their dedication and collaboration on this project. Each person I have encountered used his skills and talents to help me finish this project.

Finally, I’m blessed to have a family that believes in me, supportive of higher education and the future of technology. With their patience, I was able to put in the time needed to work on my educational journey. I am appreciative of this unconditional support.

**Table of Contents**

[Chapter 1: Introduction: 1](#_Toc26491665)

[1.1 Overview: 1](#_Toc26491666)

[1.2 Motivations of the project: 1](#_Toc26491667)

[1.3 Problem statment: 2](#_Toc26491668)

[1.4 Aims and objectives: 2](#_Toc26491669)

[1.5 Scope and constraints of the project: 3](#_Toc26491670)

[1.6 Suggested solution 3](#_Toc26491671)

[1.7 Project plan: 4](#_Toc26491672)

[Chapter 2: literature review 5](#_Toc26491674)

[2.1 introduction: 5](#_Toc26491676)

[2.2 overview of Hospital resources and bed occupancy 5](#_Toc26491677)

[2.3 existing solutions for bed occupancy predections: 5](#_Toc26491678)

[2.4 machine learning approaches in health-care: 6](#_Toc26491679)

[2.5 challanges and limitations of existing solutions 7](#_Toc26491680)

[2.6 Summary: 7](#_Toc26491681)

[Chapter 3: Requirements and Analysis 8](#_Toc26491682)

[3.1 **Introduction to Requirements and Analysis**: 8](#_Toc26491684)

[3.2 functional requirement: 9](#_Toc26491685)

[3.3 **Non-Functional Requirements**: 9](#_Toc26491686)

[3.4 **System Architecture**: 1](#_Toc26491687)0

[3.5 Data requirments and Analysis: 1](#_Toc26491688)1

[3.6 UML Diagrams Analysis: 1](#_Toc26491689)2

[3.6.1 Use case diagram: 1](#_Toc26491690)2

[3.6.2 Activity diagram: 1](#_Toc26491691)2

[3.7 Tools and technolgies: 1](#_Toc26491689)3

[3.8 Risk Analysis: 1](#_Toc26491689)4

[3.9 Reference: 1](#_Toc26491666)4

Chapter 1: Introduction

# 1.1 Overview

The unpredictable nature of patient demand creates severe challenges in hospital bed occupancy, leading to overcrowded conditions and wastage of resources and reducing the quality of care. Effective management of available resources is a must to make hospital operations most efficient and improve patient outcomes. The "Predictive Analytics for Hospital Bed Occupancy Management" project is set to create a predictive analytics system to forecast expected bed occupancy using historical patient data; an engaged dashboard would complement this for the efficiency, transparency, and accountability of hospital administrators. The importance of this solution, to streamline resource allocation, reduce waiting time for patients, and enhance the efficiency of hospitals, cannot be overstated. Knowing the occupancy rates in advance will allow hospitals to make wise decisions on resource allocation regarding staff and equipment, thus enhancing patient care while lowering costs.

### ****1.2 Motivations of the Project****

Efficient resource management in healthcare facilities, particularly hospitals, is a growing concern worldwide. During peak periods, hospitals often face the challenge of overcrowded wards and strained resources. Conversely, periods of low demand may lead to underutilized resources. This imbalance directly impacts patient care and increases operational costs. The motivation behind this project is to leverage predictive analytics to provide hospital administrators with actionable insights, enabling them to better plan and allocate resources. With recent global health crises highlighting the importance of effective resource utilization, this project addresses a pressing societal need.

### ****1.3 Problem Statement****

Hospitals face significant challenges in managing bed occupancy rates efficiently in the context of the unpredictable admissions. This unpredictability can lead to critical situations where hospital resources are either stretched too thin or not utilized enough. For example:  overcrowding can lead to increased waiting time for patients which will delay the emergency medical care and decrease patient satisfaction. It also puts a heavy burden on medical staff, increasing the levels of stress and the risk of clinical errors. On the other side, under-utilization of hospital resources, such as empty beds, has a wasteful effect that yields needless operational costs, in the end compromising the financial survival of medical companies. This continuous mismatch between patient demand and resource availability has a direct impact on the quality and effectiveness of hospital functioning. Existing ways of administering hospital resource allocation usually rely on manual or simple statistical models, and may not be adequate to represent the intricacy of patient load arrivals or present actionable information to a hospital administrator. A urgent need exists for a reliable and powerful predictive model to be able to process historical admission data, to be able to identify trends, and to predict bed occupancy in the future. This solution would enable hospital administrators to redeploy resources in ways that could improve efficiency, minimize waste, and improve patient outcomes. Through the application of predictive analytics to hospital processes, healthcare institutions can enhance planning, optimize operations, and provide high-quality patient care.

### ****1.4 Aims and Objectives****

* **Aim**: To develop a predictive model and dashboard for hospital bed occupancy management.
* **Objectives**:

1. To develop a machine learning-based predictive model that forecasts future hospital bed occupancy using historical patient admission data.
2. To analyze key factors such as patient demographics, admission trends, and seasonal variations that impact bed occupancy rates.
3. To design a web-based dashboard that visualizes predicted bed occupancy, allowing hospital administrators to make data-driven decisions for resource allocation.
4. To improve hospital operational efficiency by reducing bed shortages or under-utilization through accurate, real-time predictions.
5. To provide health-care providers with a tool to optimize resource planning, ultimately improving the patient experience and quality of care.

### ****1.5 Scope and Constraints of the Project****

### In this project, the prediction of hospital bed occupancy would be pursued based on historical patient admission records. Data pre-processing, model development, and a dashboard for presenting predictions will be included. Nevertheless, the project can be carried out using public or simulated data, because of privacy limitation of real hospital data. Further, the scope of the project does not include integration of real-time data or sophisticated dynamic resource management mechanisms other than bed loading. ****1.6 Suggested Solution****

The proposed solution is doing a predictive analytics system which is designed to forecast future hospital bed occupancy rates using old patient admission data. This system will give the edge to machine learning algorithms to identify the patterns and trends within the data, providing accurate and actionable predictions for hospital administrators. By enabling better anticipation of bed demand, the system aims to address both overcrowding and underutilization of resources.The solution will be developed using **Python**, a widely adopted programming language in data science and machine learning. Key libraries and frameworks will include:

* **Pandas**: To pre-process and manipulate large data sets efficiently, handling tasks like cleaning, filtering, and organizing data for analysis.
* **Scikit-learn**: For building and training predictive machine learning models. This library offers robust tools for regression, classification, and evaluation metrics that are essential for forecasting bed occupancy.
* **Streamlit**: To create an intuitive and interactive dashboard for hospital administrators. The dashboard will allow users to upload data, view predictions, and interact with visualizations such as trends, graphs, and tables.

The system is a very intuitive dashboard which will make sure that hospital staff will be able to easily understand the predictions and the insights without the need for technical skills. The dashboard will also provide a current visualization of trend in past data, that the administrators will be able to contrast with model outcomes from the past to increase the quality of the decision-making process. Furthermore, in order to guarantee the accuracy and reliability of the predictions, the system will also include data validation and preprocessing stages.   
  
This solution not only effectively allocates resources, but also can increases the efficiency of the whole hospital operations. Through the reduction of bed shortages and operational inefficiencies, the system contributes to better patient care and patient satisfaction. Finally, this predictive analytics instrument fills the gap between the conventional manual planning and the contemporary data-driven decision-making in clinical practices.

### ****1.7 Project Plan****

The project will be completed in six key phases:

1. **Project Planning and Setup**: Define objectives, gather data, and set up tools.
2. **Data Collection and Preparation**: Obtain and pre-process historical hospital data.
3. **Build and Train Machine Learning Models**: Develop a predictive model to forecast bed occupancy.
4. **Build a Dashboard**: Create a web-based dashboard to display predictions.
5. **Final Testing and Deployment**: Validate the model and deploy the dashboard.
6. **Submission and Presentation**: Prepare a final report and present findings.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Task description | Start Week | Duration(Weeks) | state |
| 1 | Searching for project ideas | 1 | 1 | Done |
| 2 | Writing and submitting proposal | 1 | 1 | Done |
| 3 | Define project borders and stakeholders | 1 | 1 | Done |
| 4 | Getting ready to start the report | 4 | 1 | Done |
| 5 | Prepare for TMA part 1 | 4 | 1 | Done |
| 6 | Define requirements (FR,NFR) | 4 | 2 | Done |
| 7 | Exams and submissions | 6 | 2 | Done |
| 8 | Draw all needed diagrams | 8 | 1 | Done |
| 9 | Submit TMA PART1 | 8 | 1 | Done |
| 10 | Prepare for report part1 | 8 | 1 | Done |
| 11 | Elicit for more information | 8 | 1 | Done |
| 12 | Submit the report | 9 | 1 | In-progress |
| 13 | Prepare for presentation | 9 | 1 | -- |
| 14 | Final exams | 10 | 3 | -- |
| 15 | Summer break | 13 | 2 | -- |
| 16 | Research data and tools | 15 | 2 | -- |
| 17 | Obtain data-sets | 17 | 3 | -- |
| 18 | Clean and reprocess data | 17 | 3 | -- |
| 19 | Experiment with algorithms | 20 | 2 | -- |
| 20 | Train and evaluate models | 22 | 4 | -- |
| 21 | Design UI/UX | 26 | 1 | -- |
| 22 | Integration predictions and visualizations | 27 | 3 | -- |
| 23 | Test model accuracy | 30 | 2 | -- |
| 24 | Deploy dashboard | 32 | 2 | -- |
| 25 | Prepare for final Presentation and finishing | 34 | 1 | -- |

**Chapter 2 : Literature Review**

#### ****2.1 Introduction****

#### This chapter gives an overview of available methods and technologies for the hospital bed occupancy prediction and the resource management optimization. The review seeks to pinpoint the advantages and disadvantages of these techniques, identify the limitations of existing solutions, and argue for the creation of a predictive model and dash board for resource management in hospitals. ****2.2 Overview of Hospital Resource Management and Bed Occupancy****

#### Hospital resource management, especially hospital bed occupancy, is fundamental to timely and appropriate patient care. Inefficient bed use may result in congestion, longer wait times for patients and health-care providers, and an overwhelmed health-care workforce. On the other hand, under-hospital-bed-capacity leads to resource inefficiency and financial losses. Successful management of bed occupancy relies on the correct prediction of patient demand and the efficient allocation of resources. ****2.3 Existing Solutions for Bed Occupancy Prediction****

#### Methods to estimate hospital beds occupancy rate have been developed. Methods based on statistics, including time-series forecasting, linear regression, and Markov models, are traditional. These methods analyze historical data to estimate future demand. More recently, machine learning (ML) models have been applied to the problem and have been found to provide higher accuracy and flexibility through learning of such complex patterns from large data sets. Decision trees, support vector machine, and neural network models are some popular ML techniques used there. One of the commercial solutions is to include data analytics platforms for the use of hospital administrators, but they do not offer a lot as far as customization, real-time prediction, and usability for smaller healthcare centers ****2.4 Machine Learning Approaches in Health-care Analytics****

Machine learning is an important tool in health care analysis, and it offers new approaches to difficult issues in disease diagnosis, personalized therapy and resource optimization. With the analysis of big data, machine learning (ML) models can detect patterns and trends that it would be hard for humans to detect, which healthcare professionals use in order to make decisions based on data.  
  
In the context of hospital resource management, machine learning plays a pivotal role in predicting patient admissions and optimizing bed occupancy. Predictive models trained on historical patient admission data can uncover hidden correlations between variables such as admission dates, patient demographics, seasonal trends, and external factors like disease outbreaks. These insights enable hospital administrators to better plan resource allocation, improve operational efficiency, and reduce costs.

Several machine learning techniques have been applied in healthcare analytics, each suited to different types of problems:

* **Supervised Learning**: Methods like random forests, decision trees, and gradient boosting are widely applied for hospital admission forecasting and resource demand prediction. That is, these algorithms train on labeled data, extracting relationships between an input feature and the target variable, such as bed occupancy rate.
* **Unsupervised Learning:** Clustering approaches, e.g., k-means or hierarchical clustering, are applied to cluster the patients on the basis of similarities of admission data or medical history. This contributes to the detection of patterns that can be used to inform hospital planning and resource specification.
* **Deep Learning**: Neural networks, and in particular, the recurrent neural networks (RNNs) and the long short-term memory (LSTM) networks are extremely powerful in the context of time-series data, e.g., patterns of patient admission trend across time. These models are strong at learning sequential dependencies, but also strong at performing accurate long-term prediction.

The availability of frameworks like **scikit-learn**, **TensorFlow**, and **PyTorch** has significantly simplified the implementation of machine learning models in health-care. These libraries provide robust tools for data preprocessing, model training, and evaluation, allowing rapid prototyping and deployment of tailored predictive systems. For example, scikit-learn's user-friendly interface supports various algorithms, making it an ideal choice for building initial models to forecast hospital bed occupancy.  
  
Although promising, the application of machine learning to health-care analytic s has its own set of challenges such as data privacy concerns, data quality issues, and model interpretability. Nevertheless, with the development of secure computing, interpreable AI, and open source machine learning pipelines, these challenges start to be overcome, and thus opening the door to a more general use.  
  
Utilizing machine learning techniques, health-care institutions are able to make effective decision making, improve patient care and make better use of resources. For hospital bed occupancy management, these methods offer practical guidance to optimize the balance between patient demand and resource availability towards better health-care systems.  
 **2.5 Challenges and Limitations of Existing Solutions**

#### Although standard statistical techniques serve as a measure for prediction, they have a limitation in their capability to deal with complex non-linear structure and data variation. ML models are more flexible for use, but also constrained by the limited availability, quality, and interpret-ability of data . Real-time integration and intuitive visualizations are additionally absent in most current solutions, further constraining their ability to be used in practice within the hospital environment. Additionally, privacy issues and data access limitations, in particular, restrict hospitals from sharing and using relevant amounts of patient data that may affect model performance and sustainability .****2.6 Summary****

Machine learning has emerged as a promising alternative, at the very least, for conducting large-scale data analyses, looking for hidden patterns, and making accurate predictions. Approaches such as random forests, gradient boosters, and neural networks have found applications in healthcare analytics for various tasks such as patient admissions forecasting and resource allocation. They efficiently deal with high-dimensional data and deploy predictions with a greater degree of precision and scalable performance. However, adverse impediments, mainly during deployment, pose restraining notions that cause added hesitancy to the many present-day machine-learning applications within the context of health applications.

Furthermore, existing systems frequently lack accessibility or user-friendliness for non-technical stakeholders, like hospital administrators, who are the primary end-users of such tools. The ideal tools must not only provide predictive power, support real-time decision-making, and communicate insights in an actionable and intuitive format

Thus, the project intends to fill this gap through a fusion of machine learning techniques with a web-based dashboard developed with the end-user in mind. It is going to use particular machine-learning tools like scikit-learn for predictive modeling and Streamlit for interactive data visualization. The predicted system will be user-friendly and practical for hospital administrators, showing forecasts of bed occupation trends in an accessible way. With this fusion, it is bound to create a solution that is accurate, adaptive, and ergonomic, consequently getting the philosophies of resource optimization returned to the optimum of health personnel.

Such an endeavor appears fairly ambitious with prospective results, given the gap cause-by present methods, controlling for resounding performance gains for hospital resource management. Such technologies offered promise by bridging developments in predictive science with the thrust of real-world requirements facing healthcare facilities in imbedding algorithms inside the decision cycle purposed toward smarter, data-driven decisions occurring within hospitals.   
  
 **Chapter 3 : Requirements and analysis**

#### ****3.1 Introduction****

This chapter offers a thorough review of the elements needed for the realization of the predictive analytics system to manage hospital bed occupancy. It starts by characterizing the functional and non-functional requirements that set the desired functionality and performance criteria for the system. Functional requirements indicate the features and functionalities that the system will do, such as bed occupancy prediction and visualization generation, while non-functional requirements span aspects of system design and usability with emphasis on reliability and performance.  
  
Additionally, this chapter delves into the system architecture, presenting a high-level design that illustrates the relationships between the system's components. Understanding the architecture is essential to ensure that the development process aligns with the intended functionality and scalability.

The chapter also on the concept of data requirements and they play an important role in training the machine learning model and providing reliable prediction. It contains a characterisation of the kind, quality, and format of the data required, as well as a description of data flow in the system. Flowcharts and use case diagrams are both provided to guide a view of the relations between the system and the user.  
  
Finally, the chapter describes the tools and technologies to be used during the development, including Python, pandas, scikit-learn and Streamlit. These tools are chosen because they are able to facilitate data preprocessing, model building, and the generation of an interactive dashboard. Based on and through the careful documentation and analysis of these requirements and system components, this chapter provides a solid base upon which to build successful implementation of the predictive analytics system.

#### ****3.2 Functional Requirements****

The functional requirements specify the features and capabilities that the system must provide:

* The system shall predict hospital bed occupancy rates based on historical data inputs.
* The system shall allow users to upload or input historical patient admission data.
* The system shall display predictions and trends on a web-based dashboard.
* The system shall visualize data in the form of charts, graphs, and tables.
* The system shall provide basic data pre-processing functionalities (e.g., handling missing values).
* The system shall offer options for users to view historical data and analyze trends.

#### ****3.3 Non-Functional Requirements****

The non-functional requirements define the characteristics of the system, including its performance, usability, and security aspects:

* **Performance**: Predictions shall be generated within 5 seconds of data input.
* **Usability**: The dashboard interface shall be user-friendly, with intuitive navigation and interactive visualizations.
* **Reliability**: The system must provide consistent and accurate predictions based on valid input data.
* **Scalability**: The system should be able to handle larger datasets with minimal performance degradation.
* **Security**: The system will ensure data security by restricting access to authorized users and protecting sensitive data.

#### ****3.4 System Architecture****

The system architecture is built to facilitate the ease of development and deployment of the predictive analytics system (that is, hospital bed occupancy management). It consists of three key elements, each of which is critical to system goals:  
  
1- **Data Collection and Pre-Processing Module**:  
This module is the basis of the whole system, it is in charge of the raw input data. It makes it possible to upload historical patient admission data sets from different formats, like the CSV or Excel format. It guarantees that data is cleaned, pre-processed, and formatted in the correct way for analysis. Tasks in this module include:

* Handling missing values by applying imputation techniques or removing incomplete records.
* Converting categorical data (e.g., admission types) into numerical formats for compatibility with machine learning algorithms.
* Normalizing and standardizing numerical features to improve model performance. This module is implemented using Python libraries like **pandas** and **NumPy**, which are well-suited for efficient data manipulation and preparation.

1. **Machine Learning Model**:  
   At its heart, the system consists of a predictive machine learning model, which is trained to predict hospital bed occupancy trends. The model is based on supervised learning, and the training is implemented with the algorithms of the **scikit-learn** library. Key features include:

* Training the model on historical patient admission data to identify patterns and trends.
* Predicting bed occupancy for specific time frames (e.g., daily or weekly).
* Evaluating model accuracy using metrics such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE). The model is modular, allowing for experimentation with different algorithms (e.g., random forests or gradient boosting) to optimize performance

**3- Web-Based Dashboard**:  
The dashboard provides an intuitive and interactive interface for hospital administrators, making the system’s insights easily accessible. Built using Streamlit, the dashboard offers the following functionalities:

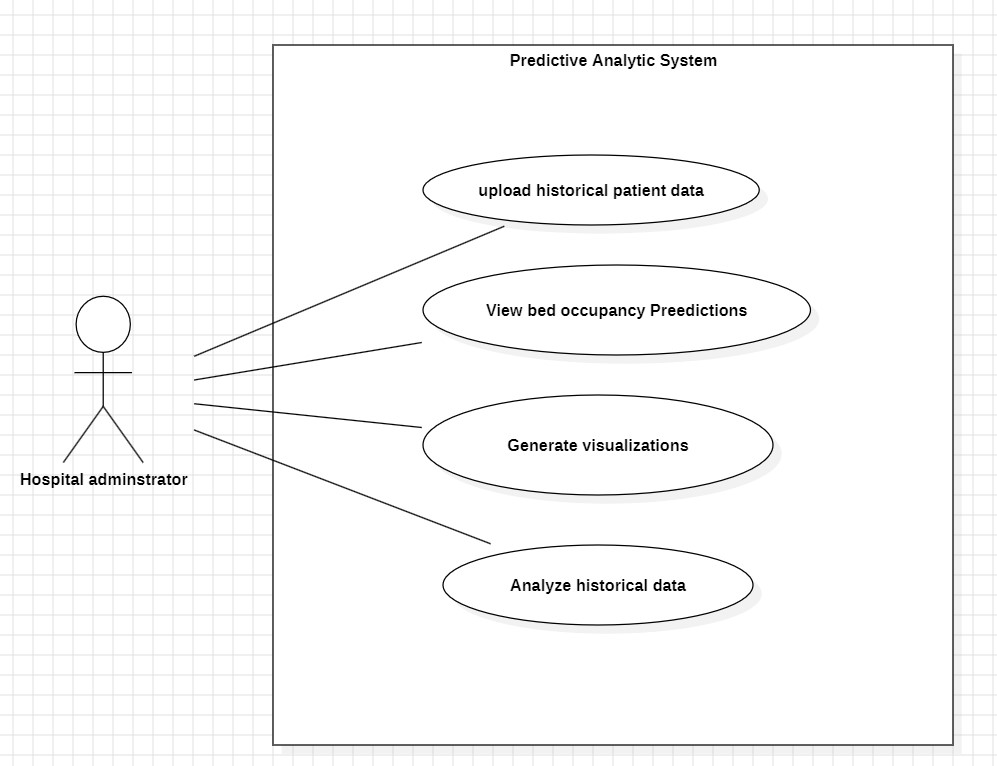
* Uploading and visualizing historical data directly within the platform.
* Displaying predictive results in an easy-to-understand format, including charts, graphs, and tables.
* Allowing users to analyze trends and make data-driven decisions. The dashboard ensures that even non-technical users can interact with the system effectively, bridging the gap between complex machine learning outputs and actionable insights.

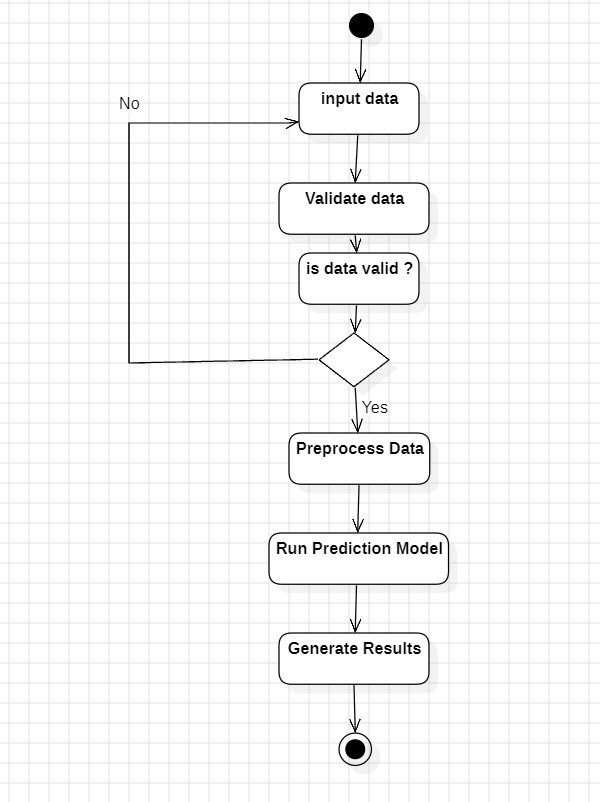
Together, these components form a cohesive architecture, enabling the system to process raw data, generate accurate predictions, and present results in a user-friendly manner. This modular design ensures scalability, making it possible to integrate additional features or handle larger datasets in the future.

#### ****3.5 Data Requirements and Analysis****

The data required for this project includes:

* **Historical Patient Admission Data**: Admission dates, discharge dates, patient demographics, and reason for admission (if available).
* **Data Analysis**: The data will be preprocessed using tools like pandas to clean and transform it for model training. The data will be analyzed to identify trends, patterns, and outliers that impact bed occupancy rates.  
    
    
    
  **3.6 UML Diagrams Analysis**

3.6.1 use case diagram   
  
  
  
  
  
  
  
3.6.2 activity diagram



#### ****3.7 Tools and Technologies****

The system will be developed using the following tools and technologies:

1. **Programming Language**: Python
2. **Libraries**:

* **pandas** for data manipulation and preprocessing
* **scikit-learn** for building and training the predictive model
* **matplotlib** and **seaborn** for data visualization
* **Streamlit** for developing the web-based dashboard

1. **Development Environment**: Jupyter Notebook and Visual Studio Code (or other Python IDEs)
2. **Version Control**: Git for tracking changes and collaborating

#### ****3.8 Risk Analysis****

Identifying potential risks ensures that we can proactively plan to mitigate them:

1. **Data Availability**: Limited access to high-quality hospital data may impact model accuracy. Mitigation: Use public datasets or generate synthetic data
2. **Model Accuracy**: The predictive model may produce inaccurate forecasts due to data variability. Mitigation: Test various algorithms and tune parameters for improved performance.
3. **Privacy Concerns**: Handling patient data requires compliance with data privacy regulations. Mitigation: Use anonymized or simulated data to avoid sensitive information breaches.

#### ****3.9 References****

1- "Machine Learning for Health-care: Trends and Challenges." HealthTechOnline. Accessed from <https://www.healthtechonline.org> on January 1, 2024.  
  
2-Kaggle. (2023). Hospital Admissions Dataset. Retrieved from <https://www.kaggle.com>.  
  
3-Pandas Developers. (2023). Pandas Documentation. Retrieved from <https://pandas.pydata.org.>  
4-Database System: The complete book , Jeffery Ullman 2001  
  
5- fundamentals of data base systems , Ramez el masri 2001