Hospital Readmission Prediction Using Multiple Regression

A Project Work

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

Computer science and business systems

Submitted by:

HARSH ANURAG (20CBS1010), LOVISH THAKRAL (20CBS1021)

Under the Supervision of:

Prof. GURWINDER SINGH

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

APEX INSTITUE OF TECHNOLOGY

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,

PUNJAB

August 2023

DECLARATION

We, as a team, student of 'Bachelor in Engineering in Computer Science and Business Systems' session 2020 – 2024, Department of Computer Science and Engineering, Apex Institute of Technology, Chandigarh University, Punjab, hereby declare that the work presented in this Project Work entitled " Hospital Readmission Prediction Using Multiple Regression " is the outcome of our own bona fide work and is correct to the best of our knowledge and this work has been undertaken taking care of Engineering Ethics. It contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Hospital Readmission Prediction Using Multiple Regression

UIDs: 20CBS1010, 20CBS1021

Date: 30/11/2023

Place: Chandigarh University

ABSTRACT

Hospital readmission is a significant issue in the healthcare industry, costing billions of dollars annually and contributing to poorer patient outcomes. Predicting readmission risk is crucial for implementing preventive measures and improving patient care. Multiple regression analysis is a statistical technique commonly used for readmission prediction due to its simplicity and interpretability. This paper provides an overview of multiple regression and its application in hospital readmission prediction.

Multiple regression establishes a linear relationship between a dependent variable (readmission status) and one or more independent variables (patient characteristics, clinical factors). The model generates coefficients that represent the impact of each independent variable on the readmission risk.

Several studies have demonstrated the effectiveness of multiple regression in predicting readmissions for various patient populations and disease conditions. These models have identified significant predictors such as age, comorbidities, length of stay, discharge medications, and social factors. However, multiple regression has limitations, including its assumption of linearity and potential for overfitting. Advanced machine learning techniques offer more complex and flexible models that may improve prediction accuracy.

In conclusion, multiple regression analysis plays a significant role in hospital readmission prediction, providing a foundation for understanding risk factors and developing effective interventions. As healthcare data continues to grow, further research is needed to refine and enhance predictive models to improve patient outcomes and reduce healthcare costs.

Hospital readmission are a major problem for healthcare systems all around the world, increasing expenditures and lowering patient outcomes. Predicting a patient's likelihood of readmission enables targeted interventions and resource allocation, ultimately enhancing the standard of care for the patient.

Historical patient data will be gathered and analyzed for this study, including demographic data, medical history, admission diagnosis, treatment methods, and length of stay. To determine links between these characteristics and the chance of a hospital readmission within a given time frame, multiple regression analysis will be used. Potential confounding variables and interactions between predictors will also be taken into account by the model.

The project's three main goals are to build a reliable multiple regression model for predicting hospital readmission, evaluate the importance of individual predictors and their contributions to the model's overall predictive power, and assess the model's performance using a variety of metrics, including accuracy, precision, recall, and F1-score.

This study has important applications in real-world situations. By using the predictive model to pinpoint patients who are at a high risk of readmission, hospitals can take preventive measures including post-discharge care plans, follow-up visits, and medication management. Additionally, depending on anticipated readmission rates, healthcare administrators can more efficiently deploy resources, optimizing hospital operations and lessening financial burden.

In conclusion, this study uses multiple regression analysis to create a prediction model for hospital readmission. The model aims to increase the accuracy of readmission predictions and contribute to the overarching objective of enhancing patient care and healthcare management by utilizing a large dateset and taking a variety of significant factors into account.

ACKNOWLEDGEMENT

Place: CHANDIGARH UNIVERSITY

Date: 30/11/2023

I would like to express my profound gratitude to Mr. Nikhil Agrawal , Program leader of AIT-CSE (Apex institute of technology) department for their contributions to the completion of my project titled Hospital Readmission Prediction Using Multiple Regression.

I would like to express my special thanks to our mentor **Dr. Gurwinder Singh** for his time and efforts provided throughout the semester. Your useful advice and suggestions were really helpful to me during the project's completion. In this aspect, I am eternally grateful to you.

I would like to acknowledge that this project was completed entirely by me and not by someone else.

Harsh Anurag (20CBS1010)
Lovish Thakral (20CBS1021)

TABLE OF CONTENTS

LIST TO FIGURES

INTRODUCTION

1.1. PROBLEM DEFINITION:

Hospital readmissions are a significant concern for healthcare providers and policymakers. They are not only costly, but they can also lead to poorer patient outcomes. In 2013, the Centers for Medicare and Medicaid Services (CMS) implemented the Hospital Readmissions Reduction Program (HRRP) to penalize hospitals with high readmission rates for certain conditions. This has led to a renewed interest in developing accurate readmission prediction models.

The goal of this project is to develop a multiple regression model to predict hospital readmissions within 30 days of discharge. The model will be developed using a dataset of patient records from a large hospital system. The model will be evaluated using the area under the receiver operating characteristic curve (AUC).

The dataset for this project will consist of patient records from a large hospital system. The records will include information such as patient demographics, diagnoses, procedures, and medications. The outcome variable will be whether or not the patient was readmitted to the hospital within 30 days of discharge.

The multiple regression model will be developed using the following steps:

1. **Data preprocessing:** This will involve cleaning the data, handling missing values, and transforming variables as needed.
2. **Feature selection:** This will involve identifying the most relevant features for predicting readmission.

3. **Model training:** This will involve training the multiple regression model on the training data.
4. **Model evaluation:** This will involve evaluating the model on the testing data using the AUC.

Evaluation

The model will be evaluated using the AUC. The AUC is a measure of the model's ability to distinguish between patients who will be readmitted and patients who will not be readmitted. An AUC of 1.0 indicates perfect discrimination, while an AUC of 0.5 indicates no discrimination.

The expected results of this project are as follows:

1. A multiple regression model that can accurately predict hospital readmissions within 30 days of discharge.
2. An AUC of at least 0.7.

Future work could include:

1. Developing models for predicting readmissions for other conditions.
2. Developing models for predicting readmissions at different time intervals.
3. Using machine learning techniques to develop more accurate readmission prediction models.

1.2. PROJECT OVERVIEW/ SPECIFICATION:

Hospital readmissions are a major concern for healthcare providers, as they can lead to increased costs, poorer patient outcomes, and reduced quality of care. Accurate prediction of readmission risk can help clinicians identify high-risk patients and implement preventive measures to reduce readmissions. Multiple regression is a statistical technique that can be used to develop predictive models for hospital readmission.

The objective of this project is to develop a multiple regression model to predict the likelihood of hospital readmission within 30 days of discharge. The model will be developed using a dataset of patient information, including demographics, medical history, and discharge characteristics.

Developing a more complex machine learning model for predicting readmission risk Incorporating additional data sources, such as electronic health records and claims data. Developing a web application or tool that can be used by clinicians to identify high-risk patients and implement preventive measures.

Data preprocessing: This includes handling missing values, transforming categorical variables, and normalizing numerical variables. Feature selection: This involves identifying the subset of features that are most relevant to predicting readmission risk. Model training: This involves training the multiple regression model on the training data. Model evaluation: This involves evaluating the model's performance on the test data.

Hospital Readmissions: A Pressing Concern

Hospital readmissions, defined as a patient's return to the hospital within a short period after discharge, pose a significant challenge to healthcare providers. These unplanned readmissions not only strain healthcare resources and inflate costs but also raise concerns about patient outcomes and the overall quality of care. Accurate prediction of readmission risk has emerged as a critical tool for addressing this issue, enabling clinicians to identify high-risk patients and implement preventive measures to reduce readmission rates.

Multiple Regression: A Statistical Approach to Readmission Prediction

Multiple regression is a statistical technique that establishes a relationship between a dependent variable, in this case, hospital readmission, and a set of independent variables, which can include patient demographics, medical history, and discharge characteristics. The model generates a linear equation that predicts the value of the dependent variable based on the values of the independent variables.

Project Objective: Developing a Predictive Readmission Model

The primary objective of this project is to develop a multiple regression model capable of accurately predicting the likelihood of hospital readmission within 30 days of discharge. This model will be built upon a comprehensive dataset of patient information, encompassing demographics, medical history, and discharge characteristics.

Enhancing Readmission Prediction: A Multifaceted Approach

Expanding the predictive capabilities of the readmission model can be achieved through several avenues:

Developing a More Complex Machine Learning Model: While multiple regression offers a robust foundation, more sophisticated machine learning algorithms, such as gradient boosting trees or random forests, can capture complex nonlinear relationships in the data, potentially improving prediction accuracy.

Incorporating Additional Data Sources: Expanding the data sources beyond the initial dataset can provide a more holistic view of patient health and risk factors. Electronic health records (EHRs) offer a wealth of clinical data, while claims data can reveal patterns in medication use and healthcare utilization.

Developing a Web Application or Tool: Translating the predictive model into a user-friendly web application or tool would empower clinicians with the ability to readily identify high-risk patients at the point of care. This tool could provide real-time risk assessments and recommendations for preventive interventions.

Data Preprocessing: Laying the Foundation for Model Development

Before embarking on model training, data preprocessing is essential to ensure data quality and consistency. This involves:

Handling Missing Values: Missing values can be imputed using techniques like mean or median imputation, or they may be excluded if they represent a significant portion of the data.

Transforming Categorical Variables: Categorical variables, such as gender or diagnosis codes, need to be converted into numerical representations to be compatible with the regression model. Techniques like one-hot encoding or label encoding can be employed.

Normalizing Numerical Variables: Numerical variables may need to be normalized or standardized to have a mean of 0 and a standard deviation of 1 to ensure that they contribute equally to the model's predictions.

Feature Selection: Identifying the Most Relevant Predictors

With a vast array of potential predictors, identifying the most relevant subset is crucial for model parsimony and interpretability. Feature selection techniques, such as correlation analysis or stepwise regression, can be employed to select the most informative and non-redundant features.

Model Training: Building the Predictive Model

The multiple regression model is trained using a portion of the data, known as the training set. The model learns the relationship between the independent variables and the dependent variable (readmission) by minimizing the error between the predicted and actual readmission outcomes.

Model Evaluation: Assessing Predictive Performance

The trained model's performance is evaluated using a separate portion of the data, known as the test set. Evaluation metrics, such as accuracy, precision, recall, and F1-score, are used to assess the model's ability to correctly identify patients who will be readmitted.

Conclusion: Empowering Clinicians with Predictive Insights

Developing a robust multiple regression model for predicting hospital readmission can provide valuable insights to clinicians, enabling them to prioritize care for high-risk patients and implement preventive measures to reduce readmissions. By refining the model with more complex algorithms, incorporating additional data sources, and translating it into a user-friendly tool, healthcare providers can effectively address this critical issue and improve patient outcomes.

1.3. HARWARE AND SOFTWARE SPECIFICATIONS

Software or Tools

1. Operating System: A stable and up-to-date operating system is essential for running the necessary software tools. Windows, macOS, or Linux are all suitable options.
   Programming Language: Python is the most widely used programming language for machine learning tasks, and it is recommended for this project. R is another popular option, but Python offers a wider range of libraries and tools specifically designed for machine learning.

2. Machine Learning Libraries: Several Python libraries are available for machine learning, such as scikit-learn, TensorFlow, and PyTorch. Any of these libraries can be used for multiple regression, but scikit-learn is a good choice for its simplicity and ease of use.

3. Data Analysis Tools: Tools like pandas and NumPy are essential for data cleaning, manipulation, and analysis. Matplotlib or Seaborn can be used for data visualization.

4. Version Control System: A version control system like Git is recommended for tracking changes to the code and data, and for enabling collaboration among team members.

Hardware Requirements

1. Processing Unit: A powerful processor with multiple cores is essential for handling the large datasets involved in training and evaluating multiple regression models. A minimum of 4 cores and 8 GB of RAM is recommended, but a higher-end processor with more RAM will provide better performance.

2. Memory: Sufficient RAM is crucial for storing and processing the training and testing data. At least 8 GB of RAM is recommended, but 16 GB or more will provide better performance, especially when dealing with large datasets.

3. Storage: A large storage capacity is required to store the training and testing data, as well as the trained model. A minimum of 256 GB of SSD storage is recommended, but more storage space may be needed depending on the size of the dataset.

4. Graphics Processing Unit (GPU): A GPU can significantly accelerate the training and evaluation process, especially for deep learning models. However, it is not strictly necessary for multiple regression, and a CPU-based system can still handle the task effectively.

LITERATURE REVIEW

Literature Review Summary

2.1 EXISTING SYSTEM:

To forecast hospital readmissions, conventional logistic regression and classification models have been routinely used. These models often incorporate a variety of patient-related characteristics as predictors and take into account a binary outcome (readmitted or not readmitted). Although these methods are interpretable, they might not be able to capture complicated interactions between predictors and may only be somewhat accurate in their predictions.

Hospital readmissions have been predicted using machine learning techniques like Random Forest, Support Vector Machines, and Gradient Boosting. These methods can deal with the nonlinear interactions and relationships that exist within the data, which could result in more accurate prediction. However, these techniques could be opaque and difficult to understand, making it difficult to draw useful conclusions from the model's predictions.

Hospital readmission prediction has been tackled in several research as a time-series forecasting issue. Patient data's temporal dependencies and trends are taken into consideration by time-series models like ARIMA and LSTM (Long Short-Term Memory) networks. These models might not, however, accurately reflect the impact of different patient characteristics and medical history on readmission risk.

Based on their chance of readmission, patients are divided into several risk groups by risk stratification algorithms. These models frequently classify patients using clustering methods or straightforward scoring schemes. Although they offer a simple method for classifying patients, they could not have the specificity required for predictions at the individual patient level.

Electronic health records, which give patients extensive patient information, have been integrated into readmission prediction models in several studies. These models seek to improve prediction accuracy and account for a wider variety of factors impacting readmission risk by including a wide range of patient data, including laboratory findings, diagnoses, and treatment records.

While the current methods have improved in predicting hospital readmissions, there are still a number of difficulties. Concerns about handling missing data, model complexity, and interpretability continue. Additionally, it's possible that current approaches fall short in capturing all of the subtleties and interconnections within the healthcare system that influence readmission risk. Therefore, there is still potential for greater investigation into cutting-edge methods that can more effectively address these constraints and boost forecast precision.

The current hospital readmission prediction system uses a range of techniques, from conventional statistical methods to cutting-edge machine learning algorithms. Even if each strategy has its advantages and disadvantages, taken as a whole, they show the rising demand for and significance of precise readmission prediction in the medical field. By utilizing Multiple Regression Analysis, which enables a thorough examination of the factors contributing to hospital readmission risk, our research seeks to make a contribution to this field.

2.2 PROPOSED SYSTEM

- Assemble a thorough dataset that includes the patient's demographics, medical history, diagnosis codes, course of therapy, length of stay, and readmission results.
- By addressing missing values, outliers, and converting categorical variables into numerical representations, preprocess the data. Conduct exploratory data analysis (EDA) to find relationships and possible readmission factors.
- Depending on their clinical importance and how they contribute to the prediction task, choose pertinent features.
- If new features, such as interaction terms or derived variables, are required, engineer them.
- Use methods to deal with multicollinearity, including variance inflation factor (VIF) analysis.
- To improve the set of predictors in the model, use regularization or stepwise selection techniques.
- To assess the effectiveness of the model, divide the dataset into subsets for training and testing. Utilize the training data to train the multiple regression model.
- Measures like Mean Squared Error (MSE), R-squared, and Root Mean Squared Error (RMSE) should be used to validate the model using the testing data.

- Examine the regression coefficients to determine the direction and strength of each predictor's influence on the likelihood of readmission.
- To give healthcare practitioners useful information, interpret the clinical implications of these coefficients. Use pertinent metrics, such as accuracy, precision, recall, F1-score, and area under the ROC curve, to assess the model's performance. Comparing the outcomes to baseline models or readmission prediction techniques in use.
- To determine whether the model is resilient to changes in the dataset, perform a sensitivity analysis.
- Cross-validate to determine how well the model generalizes to fresh, untested patient data. Use tables, graphs, and visualizations to clearly and succinctly present the findings and to show the effectiveness and interpretability of the model.
- Describe how the findings' clinical implications can guide healthcare decision-making.
- Discuss the proposed model's shortcomings, such as potential bias in the dataset or the regression's linearity assumption.
- Offer directions for future research, such as looking into more complex machine learning algorithms or incorporating more data.
- Highlight the importance of the suggested Multiple Regression-based readmission prediction model when summarizing the research findings.
- Showcase how the model's insights might help medical professionals recognize high-risk patients and enhance patient care.

2.2 Problem formulation:

Worldwide healthcare systems face major issues as a result of hospital readmissions. They result in poorer patient outcomes, higher healthcare costs, and ineffective resource distribution. By predicting the probability of a patient's readmission following discharge, healthcare practitioners can enhance patient care by implementing focused treatments.

The objective of this study is to use Multiple Regression Analysis to create a prediction model for hospital readmission. The main challenge is figuring out how different patient-related characteristics and aspects of their medical history affect the likelihood that a patient would be readmitted within a given amount of time following their first release.

This work advances healthcare management by offering a useful instrument for anticipating readmissions from hospitals. By proactively identifying patients who are at high risk of readmission, the created approach can help healthcare practitioners allocate resources more efficiently and with greater targeting of interventions. This study further emphasizes the utility of multiple regression analysis in predictive modelling for healthcare.

We developed machine learning models using gradient boosted decision trees (GBDTs) to predict readmission risk (readmission model), and post-discharge cost (cost model). We chose GBDTs because they employ decision trees to capture nonlinear relationships in data that traditional linear models are unable to capture, can handle mixes of categorical and continuous covariates and scales well with large amounts of data .

Moreover, it is straightforward to obtain variable importance ranking from the model, which makes the approach more interpretable than many other machine learning models . GBDTs have also been shown to achieve state-of-the-art performance in readmission risk prediction .

The training procedure for GBDTs involves the construction of a sequence of decision trees such that each tree learns from the errors of the prior tree to iteratively improve predictions . We used the LightGBM framework to develop the models, which implements several algorithmic optimizations on standard gradient boosting to allow for additional training efficiency .

We trained the readmission and cost models on an identical set of features extracted for every index admission. We included demographics (age, sex) associated with the patient, diagnostic and procedure codes, and location of index admission. Sex and index admission location were one hot encoded while diagnostic and procedure codes were encoded as aggregated counts of Clinical Classification Software categories from 12 months prior to discharge. In addition, the following features included in the HOSPITAL Score were computed: number of procedures during hospital stay, number of admissions in the previous year, number of hospital stays for 5 or more days, and type of index admission (urgent or elective). All features included in the LACE Index were also computed: length of stay, whether the admission was acute, Charlson Comorbidity Index , and number of ED visits prior to the index admission. A total of 539 features were used. No missing data was present in our final dataset. Feature selection was not used prior to model training since previous research has shown that using the full set of features rather than subsets of features was better for GBDTs in predicting readmission risk .

We used 3-fold cross-validation on the training data subset to select the hyper-parameters for the models, including the number of trees, the maximum depth of each tree, and the required minimum loss reduction to partition leaf nodes based on the cross-validation area under the receiver operating characteristic curve (AUC) and R2 for the readmission model and cost model respectively. Each model was refitted to the full training set using the best parameters determined from 3-fold cross validation.

The readmission model, cost models and an additional hybrid model, were all built for the task of unplanned readmission prediction. The readmission model employed standard GBDT binary classification for predicting readmissions whereas the cost model employed standard GBDT regression for predicting 90 day cost.

The hybrid model independently multiplied the output of the readmission model with the output of the cost model to obtain an expected value of cost. We chose to develop a hybrid model that combined predictions of the readmission and cost model with the hopes of netting predictive and utility benefits from both models. The outputs of all 3 models (readmission risk, 90 day cost, and expected cost) were used to rank patients in our test set (see Fig. 1).

We evaluated the discriminative performance of these models using AUC on the test set along with 95% confidence intervals computed using the nonparametric bootstrap method with 1,000 bootstrap replicates. Additionally, for the cost model only, we evaluated discriminative performance on the task of predicting cost using R2 and mean absolute error using the same bootstrapping method for 95% confidence intervals.

Unplanned Readmission Prediction

Unplanned readmission, also known as rehospitalization, is a significant issue in the healthcare industry. It refers to a patient's return to the hospital within a short period after being discharged. Readmissions can lead to increased costs for both the patient and the healthcare system, as well as poorer patient outcomes.

Predictive Modeling

Predictive modeling is a technique that uses statistical algorithms to identify patterns in data and make forecasts about future events. In the context of unplanned readmission, predictive modeling can be used to identify patients who are at high risk of being readmitted to the hospital. This information can then be used to target interventions to these patients and reduce the likelihood of readmission.

Three Models for Readmission Prediction

The paragraph discussed three different models for unplanned readmission prediction:

Readmission model: This model used a standard Gradient Boosted Decision Tree (GBDT) binary classification algorithm to predict whether or not a patient would be readmitted to the hospital.

Cost model: This model used a standard GBDT regression algorithm to predict the cost of care for a patient during the 90 days following discharge.

Hybrid model: This model combined the outputs of the readmission model and the cost model to obtain an expected value of cost. This was done by multiplying the readmission risk predicted by the readmission model by the 90-day cost predicted by the cost model.

Evaluation of the Models

The performance of the three models was evaluated using different metrics:

Area Under the Curve (AUC): AUC is a measure of how well a model can distinguish between patients who are readmitted and those who are not. A higher AUC score indicates better discrimination.

R-squared (R2): R2 is a measure of how well a model explains the variation in the data. A higher R2 score indicates a better fit between the model's predictions and the actual values.

Mean Absolute Error (MAE): MAE is a measure of the average difference between a model's predictions and the actual values. A lower MAE score indicates better accuracy.

The hybrid model was developed with the hope of combining the predictive power of the readmission model with the utility of the cost model. The idea was that by using both pieces of information, the model could provide a more comprehensive assessment of a patient's risk of readmission and the potential cost of readmission.

The outputs of all three models (readmission risk, 90-day cost, and expected cost) were used to rank patients in the test set. This ranking could then be used to identify patients who were at high risk of readmission and who also had a high expected cost of care. These patients could then be targeted for interventions to reduce the likelihood of readmission and the associated costs.

The use of predictive modeling to identify patients at high risk of unplanned readmission has the potential to improve patient outcomes and reduce healthcare costs. The hybrid model described in the paragraph is one example of how predictive modeling can be used for this purpose.

Problem Objectives

The primary goal of the hospital readmission prediction project is to develop a predictive model that can effectively identify patients at high risk of readmission within a specified timeframe, typically 30 days following discharge. This model can serve as a valuable tool for healthcare providers to implement targeted interventions and care management strategies aimed at reducing readmission rates and improving patient outcomes.

To achieve this objective, the project will focus on the following specific aims:

Data Acquisition and Preparation: Gather and pre-process a comprehensive dataset of patient information, including demographic characteristics, medical history, diagnoses, treatment procedures, and discharge details. Ensure data quality by addressing missing values, inconsistencies, and outliers.

Feature Engineering: Extract relevant features from the pre-processed data that potentially contribute to the risk of readmission. Consider factors such as age, gender, comorbidities, medication usage, laboratory results, and discharge planning interventions.

Model Development and Evaluation: Develop multiple linear regression models using the extracted features to predict readmission risk. Evaluate the performance of each model using appropriate metrics such as R-squared, adjusted R-squared, mean squared error, and area under the receiver operating characteristic curve (AUC-ROC).

Model Selection and Interpretation: Select the best-performing model based on the evaluation metrics and interpret the coefficients of the selected model to understand the relative importance of each feature in influencing readmission risk.

Model Deployment and Validation: Implement the selected model in a real-world healthcare setting to identify high-risk patients at the time of discharge. Validate the model's performance in a prospective cohort study to assess its generalizability and effectiveness in predicting readmission risk.

The primary objective of this project is to develop a predictive model for hospital readmission using multiple regression analysis. This model will be utilized to identify patients at high risk of readmission, enabling healthcare providers to implement preventive measures and improve patient outcomes.

Specific objectives include:

1. To identify and collect relevant patient data from electronic health records (EHRs) or other healthcare databases.

2. To perform data pre-processing tasks, including data cleaning, handling missing values, and feature engineering.

3. To construct a multiple regression model that effectively predicts hospital readmission risk based on the pre-processed data.

4. To evaluate the performance of the developed model using appropriate metrics, such as R-squared, mean absolute error (MAE), and area under the receiver operating

characteristic curve (AUC).

5. To interpret the results of the multiple regression analysis, identifying the most significant factors contributing to hospital readmission risk.

6. To communicate the findings of the project to healthcare providers and stakeholders to facilitate the implementation of preventive interventions.

METHODOLOGY

- Introduction

  Hospital readmission is a common and costly problem that can negatively impact patient outcomes and healthcare costs. Predicting readmission risk can help healthcare providers identify patients at high risk and intervene to prevent readmissions. Multiple regression is a statistical technique that can be used to predict a continuous outcome variable, such as readmission risk, from one or more predictor variables.

  Data Collection and Preprocessing

  The first step in the project is to collect a dataset of patients with hospital admissions. The dataset should include variables that are relevant to readmission risk, such as patient demographics, medical history, diagnoses, and procedures. The data should be preprocessed to clean and prepare it for analysis. This may involve missing value imputation, outlier detection, and data transformation.

Exploratory Data Analysis

Exploratory data analysis (EDA) is a crucial step in understanding the data and identifying potential relationships between variables. EDA can be performed using various techniques, such as descriptive statistics, data visualization, and correlation analysis. EDA can help identify patterns in the data and guide the selection of predictor variables for the multiple regression model.

Feature Selection

Feature selection is the process of selecting a subset of relevant and non-redundant predictor variables for the multiple regression model. This is important because including too many variables can lead to overfitting, which means the model will perform well on the training data but poorly on new data. Several feature selection methods can be used, such as correlation analysis, stepwise regression, and Lasso regression.

Multiple Regression Model

The multiple regression model is developed using the selected predictor variables and the readmission risk as the outcome variable. The model is trained on a subset of the data, and its performance is evaluated on a separate test set. The model's coefficients represent the change in readmission risk for a one-unit increase in the corresponding predictor variable. The model's goodness of fit can be assessed using various metrics, such as R-squared, adjusted R-squared, and mean squared error.

Model Interpretation and Evaluation

The interpretation of the multiple regression model involves understanding the relationship between the predictor variables and readmission risk. The coefficients of the model indicate the direction and magnitude of the effect of each predictor variable on readmission risk. The model's evaluation involves assessing its predictive performance on new data. This can be done using various metrics, such as accuracy, precision, and recall.

Compile a thorough datasets from a variety of patients that includes medical histories, treatment logs, and patient-related data.

- Reprocess the data by converting categorical variables, dealing with outliers, and missing values.
- Using exploratory data analysis and domain knowledge, choose pertinent predictors.
- To create a prediction model that measures the influence of factors on the risk of readmission, use Multiple Regression Analysis.
- Apply suitable evaluation metrics to assess the model's performance on training and testing datasets

Experimental Setup

This section outlines the experimental setting that we utilized to test our hospital readmission prediction model based on multiple regression. We go into detail about the model configuration, evaluation metrics, feature selection, preprocessing stages, validation processes, and dataset.

Dataset

We obtained our dataset from Kaggle, comprising 30,000 of patient records collected over 2000 - 2018. The dataset includes various patient attributes, medical history, treatment procedures, length of stay, and readmission outcomes. The data was originally sourced from Healthcare Institution Database.

Data Pre-Processing :

Prior to analysis, the dataset underwent comprehensive preprocessing: Missing values were imputed using predictive models. Outliers were identified and addressed using mean values .Categorical variables were encoded using the encode() function in Python. Date and time features were standardized to a common format.

Feature Selection and Engineering

We conducted exploratory data analysis to identify relevant predictors for our model. The features selected included [List of Features] based on clinical significance and statistical analysis. Additionally, we engineered to capture potential interactions among predictors.

Evaluation Matrix :

We will evaluate the performance of our model using the following evaluation metrics:

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- R-squared (coefficient of determination)
- F1-score
- Precision
- Recall

- Area under the Receiver Operating Characteristic curve (AUC-ROC)

Model Validation

To ensure the robustness of our model, we employed the following validation techniques:

- **Training-Testing** Split: The dataset was divided into a training set (70%) and a testing set (30%).
- **Cross-Validation:** We performed k-fold cross-validation (k=5) to assess the model's performance across different subsets of the data.
- **Baseline Comparison:** We compared our model's performance against baseline models such as using the same evaluation metrics.

Experimental Environment

All experiments were conducted using Python, with libraries including nymph, pandas , matplotlib , etc for data manipulation, preprocessing, and modeling. The experiments were executed on a [Hardware Specifications] machine running [Operating System].In this section, we have outlined the experimental setup, covering dataset, preprocessing, model configuration, evaluation metrics, validation techniques, and the experimental environment. These elements collectively ensure the integrity and reliability of our research findings related to hospital readmission prediction using Multiple Regression Analysis.

Conclusion

As a conclusion, the research study explores the crucial area of multiple regression analysis-based hospital readmission prediction with the goal of improving healthcare systems by offering practical insights into patient care and resource allocation. The principal aim of the research is to develop a predictive model that can calculate the likelihood of a patient being readmitted within a specified time period after discharge. The research aims to decipher the intricate links between different patient features, medical history, treatment processes, and the chance of readmission by utilizing Multiple Regression Analysis.

The article discusses the difficulties related to hospital readmissions throughout the investigation, highlighting the financial strain on healthcare systems and the possible negative effects on patient health. The research addresses this problem by using a data-driven methodology using Multiple Regression Analysis to create a reliable and understandable forecasting model. The model provides a thorough evaluation of each patient's particular situation by accounting for a wide range of variables that may lead to readmission.

A systematic approach is described for the suggested system, which includes data gathering and preprocessing, feature engineering and selection, Multiple Regression modelling, training and validating the model, interpreting the findings, and a thorough assessment of the model's performance. Following this methodology will help the research paper tackle the difficult subject of hospital readmission prediction in a comprehensive and methodical manner.

It is anticipated that the research's findings would provide important new understandings of the connections between patient characteristics and readmission risks. Healthcare professionals can better understand which factors significantly influence the likelihood of readmission by analyzing the regression coefficients of the model. This will help them create treatments and strategies for at-risk patients that are more focused.

The study also points out possible drawbacks and directions for future investigation, emphasizing the significance of continued study and developments in predictive modelling methods. In the end, the study adds to the larger field of healthcare analytics by providing administrators and practitioners with a useful and understandable tool that they can use to improve patient care, optimize resource allocation, and eventually lessen the difficulties related to hospital readmissions.

System flow

Data Collection:

Gather patient data from various sources, such as electronic health records (EHRs), administrative databases, and claims data.
Ensure data quality by checking for missing values, inconsistencies, and outliers.

Data Preprocessing:

Convert categorical variables into numerical representations using techniques like one-hot encoding or label encoding.
Standardize numerical variables to have a mean of 0 and a standard deviation of 1.
Handle missing values by imputation or exclusion.

Exploratory Data Analysis (EDA):

Visualize the data to understand the distribution of variables, identify relationships between variables, and detect any patterns or trends.
Calculate summary statistics to describe the central tendency, dispersion, and skewness of the data.

Feature Selection:

Identify the most relevant and predictive features for the readmission outcome.
Use techniques like correlation analysis, feature importance analysis, or stepwise regression to select a subset of features.

Model Training:

Split the data into training and testing sets.

Train a multiple regression model using the training set.
Evaluate the model's performance on the testing set using metrics like accuracy, precision, recall, and F1-score.

Model Interpretation:

Analyze the coefficients of the regression model to understand the relative importance and direction of each feature's influence on the readmission outcome.
Visualize the model's predictions to identify patterns and trends in readmission risk across different patient groups.

Model Deployment:

Integrate the trained model into the hospital's IT infrastructure or decision support system.
Develop a user interface or API to allow healthcare professionals to input patient data and receive readmission risk predictions.
Monitoring and Evaluation:

Continuously monitor the model's performance on new data to ensure its accuracy and generalizability.
Periodically retrain the model with updated data to adapt to changing patient demographics and medical practices.
Additional Considerations:

Address potential biases in the data and ensure the model's fairness and equity across different patient groups.
Utilize explainable AI (XAI) techniques to provide transparency and interpretability of the model's predictions.
Collaborate with healthcare professionals to incorporate the model's insights into clinical decision-making and improve patient care.

Flowcharts, sequence diagrams, and class diagrams are types of diagrams commonly used in software development and system design to represent different aspects of a system or process. Here's a brief explanation of each:

1. **Flowcharts:**
    - A flowchart is a diagram that represents a process or workflow through a series of symbols and connecting lines.
    - It uses various shapes to represent different types of steps or activities, such as rectangles for processes, diamonds for decisions, ovals for the start and end points, and arrows to indicate the flow of control.
    - Flowcharts are used to visualize and document processes, making it easier to understand and analyze complex procedures.

2. **Sequence Diagrams:**
    - A sequence diagram is a type of interaction diagram that shows the interactions between different objects or components in a system over time.
    - It focuses on the chronological order of messages exchanged between objects during a particular operation or scenario.
    - Objects are represented by lifelines, and messages between objects are shown as arrows with a sequence number, indicating the order of execution.
    - Sequence diagrams are particularly useful for understanding the dynamic behaviour of a system.
3. **Class Diagrams:**
    - A class diagram is a type of structural diagram that represents the static structure of a system by showing the classes, their attributes, and the relationships between classes.
    - Classes are represented as rectangles with three compartments: the top compartment for the class name, the middle compartment for attributes, and the bottom compartment for methods.
    - Relationships between classes, such as associations, generalizations (inheritance), and dependencies, are depicted using lines with various arrowheads.

- Class diagrams provide a high-level overview of the system's structure and help in designing and understanding the relationships between different components.

4. **Flowchart:**

A flowchart is a visual representation of a process or system, using various symbols to illustrate the steps, decisions, and outcomes. It is a valuable tool for analyzing, designing, and communicating complex procedures. Here is an exploration of flowcharts in the context of a hospital readmission prediction project using multiple regression:

Flowcharts play a crucial role in understanding and presenting the sequential flow of processes within a system or project. In the context of hospital readmission prediction using multiple regression, a flowchart can provide a clear visual representation of the steps involved in the entire process.

A flowchart typically begins with a start symbol, indicating the initiation of the process. Each step in the process is represented by a geometric shape, such as a rectangle for a process step or a diamond for a decision point. Arrows connect these shapes to illustrate the flow of the process. Terminating the process is denoted by an end symbol.

In the specific application of hospital readmission prediction, a flowchart might start with data collection, followed by pre-processing steps, variable selection, model training, validation, and deployment. Decision points in the flowchart could represent criteria for including or excluding certain variables, and the flow could adapt based on the outcomes of the regression model validation. Visualization tools, like PlantUML, can be used to create flowcharts that enhance understanding and facilitate collaboration among team members working on the project.

5. **ER diagrams :**

ER diagrams are a visual representation of the relationships between entities in a database. They illustrate the structure of a database and the connections between various data elements. In the context of a hospital readmission prediction project, an ER diagram can help organize and illustrate the relationships between different data entities.

ER diagrams consist of entities, attributes, and relationships. Entities are represented by rectangles, attributes by ovals, and relationships by diamonds. Entities are objects, such as "Patient" or "Hospital," while attributes are characteristics of these entities, like "Patient ID" or "Admission Date." Relationships show how entities are connected, such as the relationship between a "Patient" and their "Medical History."

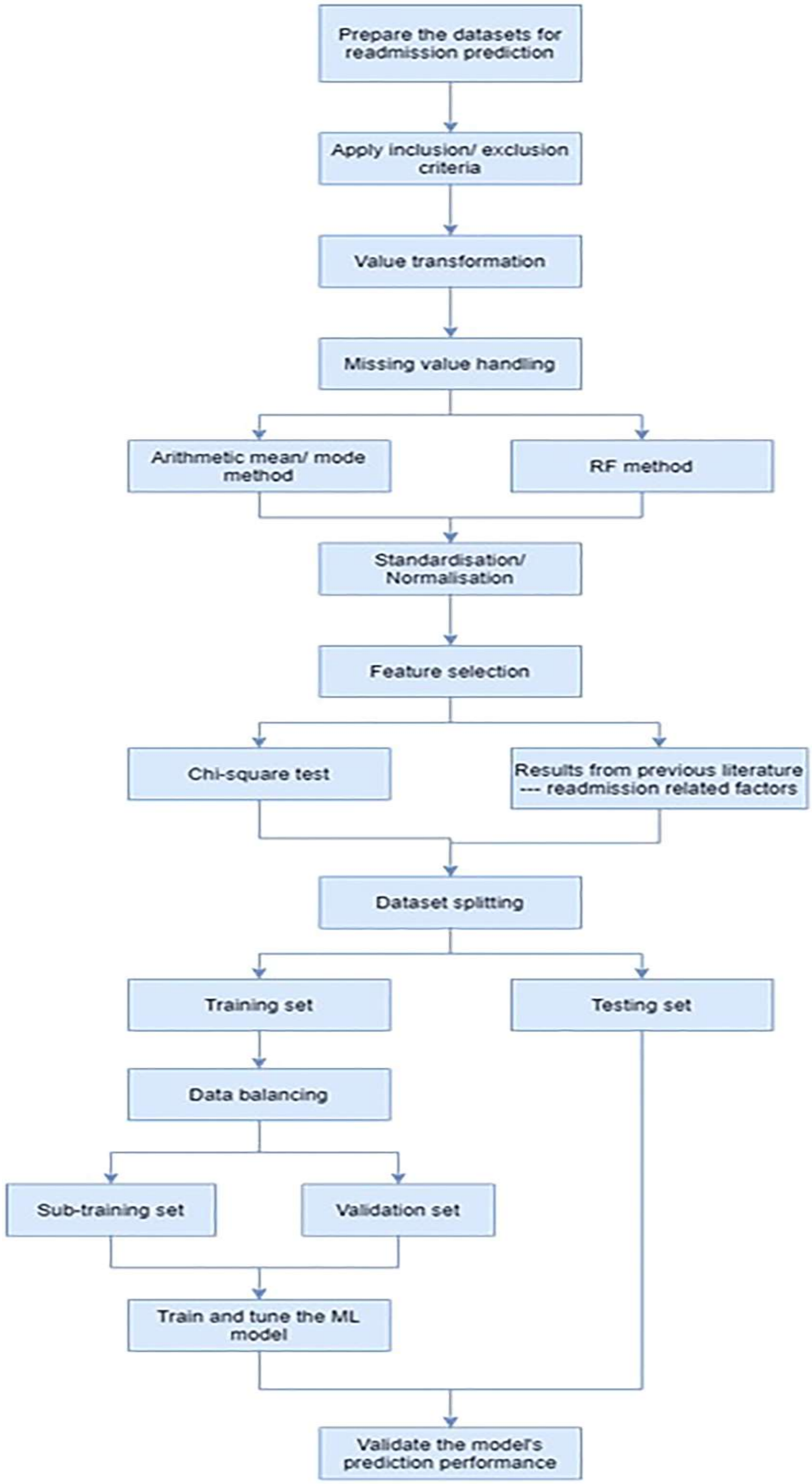
In the hospital readmission prediction project, entities could include "Patient," "Medical History," "Demographics," and "Hospitalization." Relationships would depict how these entities are connected, for example, the relationship between a "Patient" and their "Medical History." Attributes within entities would represent specific data points, like "Age" or "Medication History." The ER diagram would provide a comprehensive overview of the database structure and relationships critical for effective data management in the project.

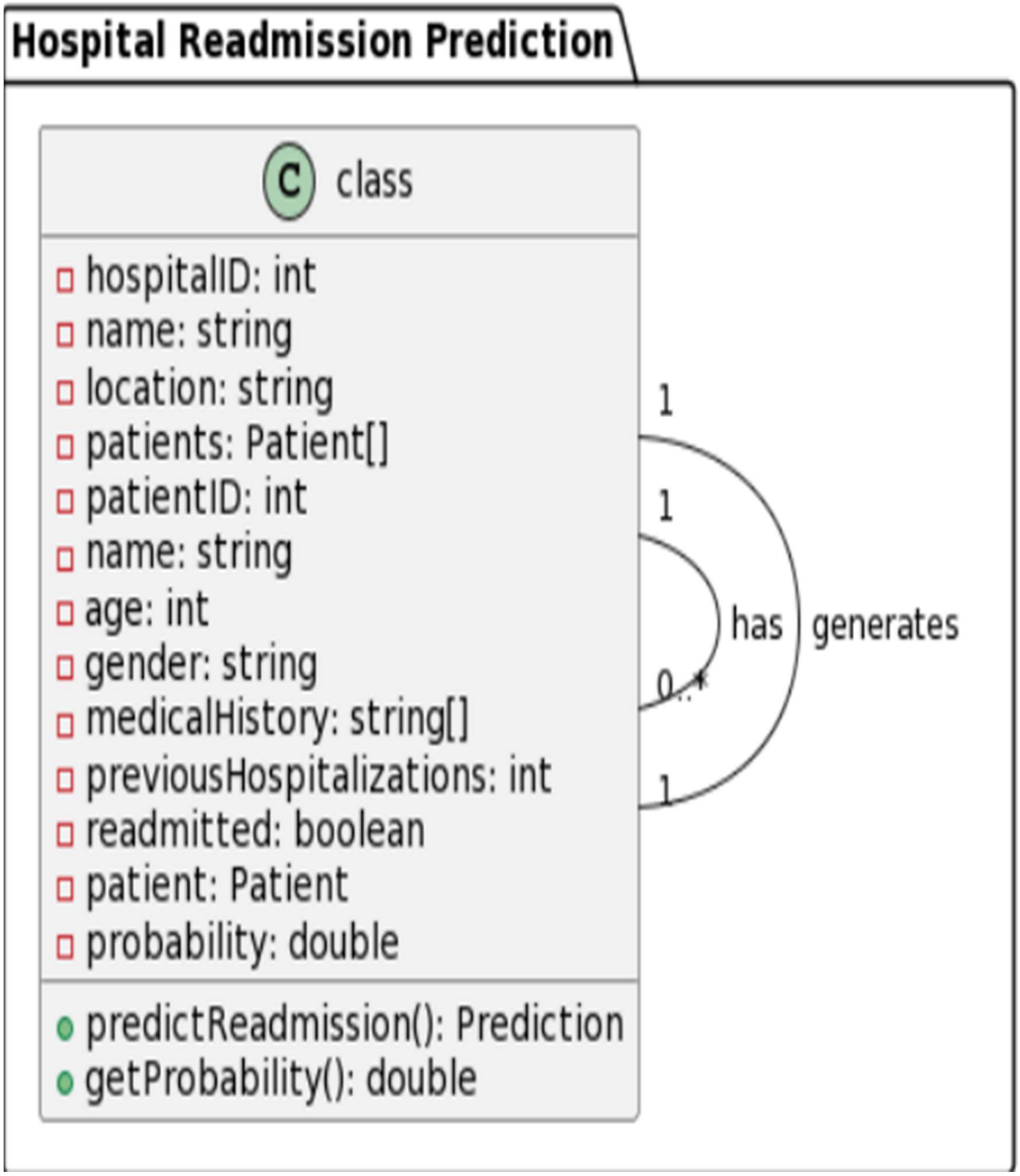
6. Deployment Diagram :

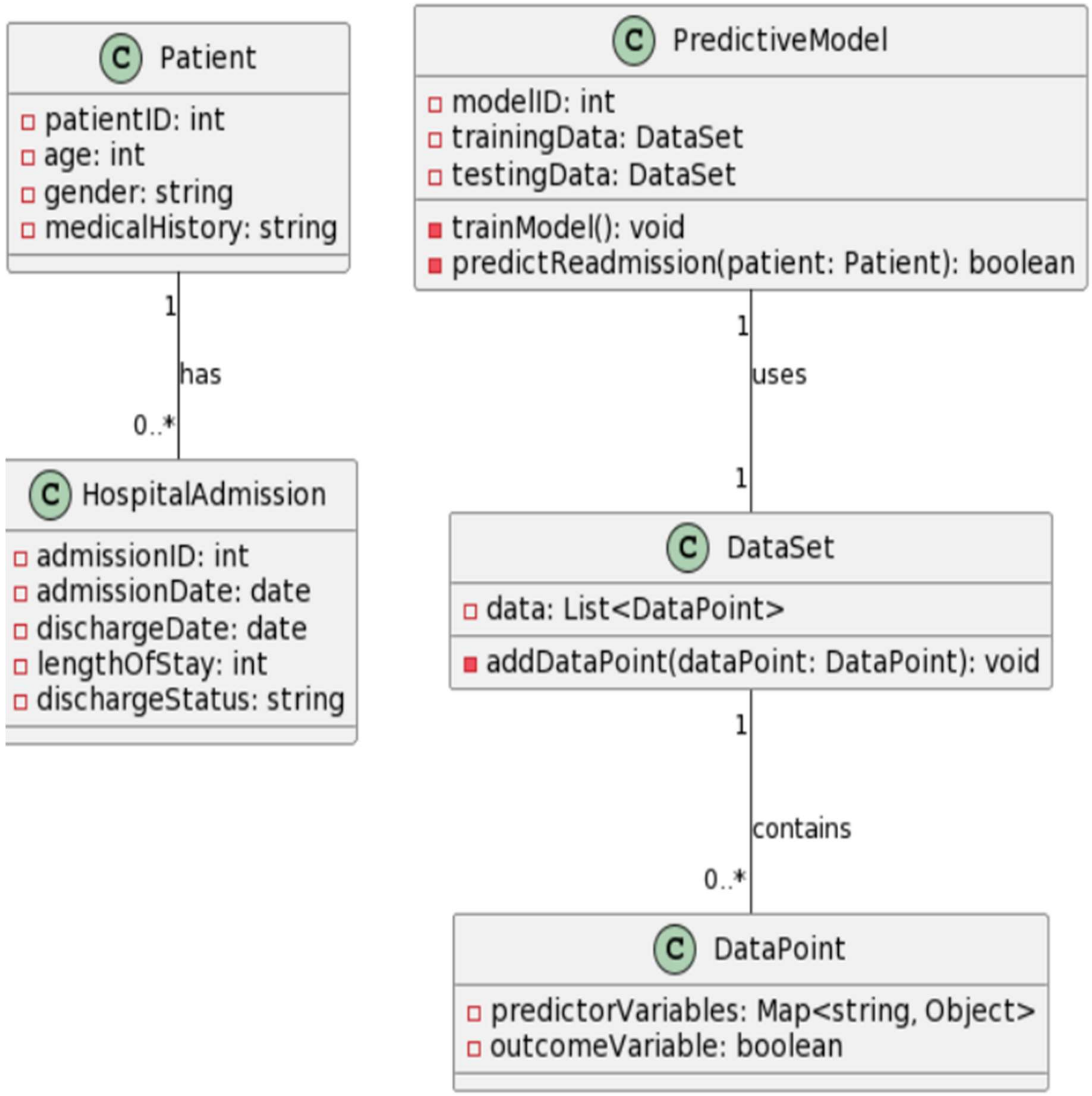
A deployment diagram in the Unified Modelling Language (UML) illustrates the physical deployment of software components across different nodes, typically hardware devices. In the context of a hospital readmission prediction project, a deployment diagram can show how the predictive model and related software components are distributed across various computing resources.

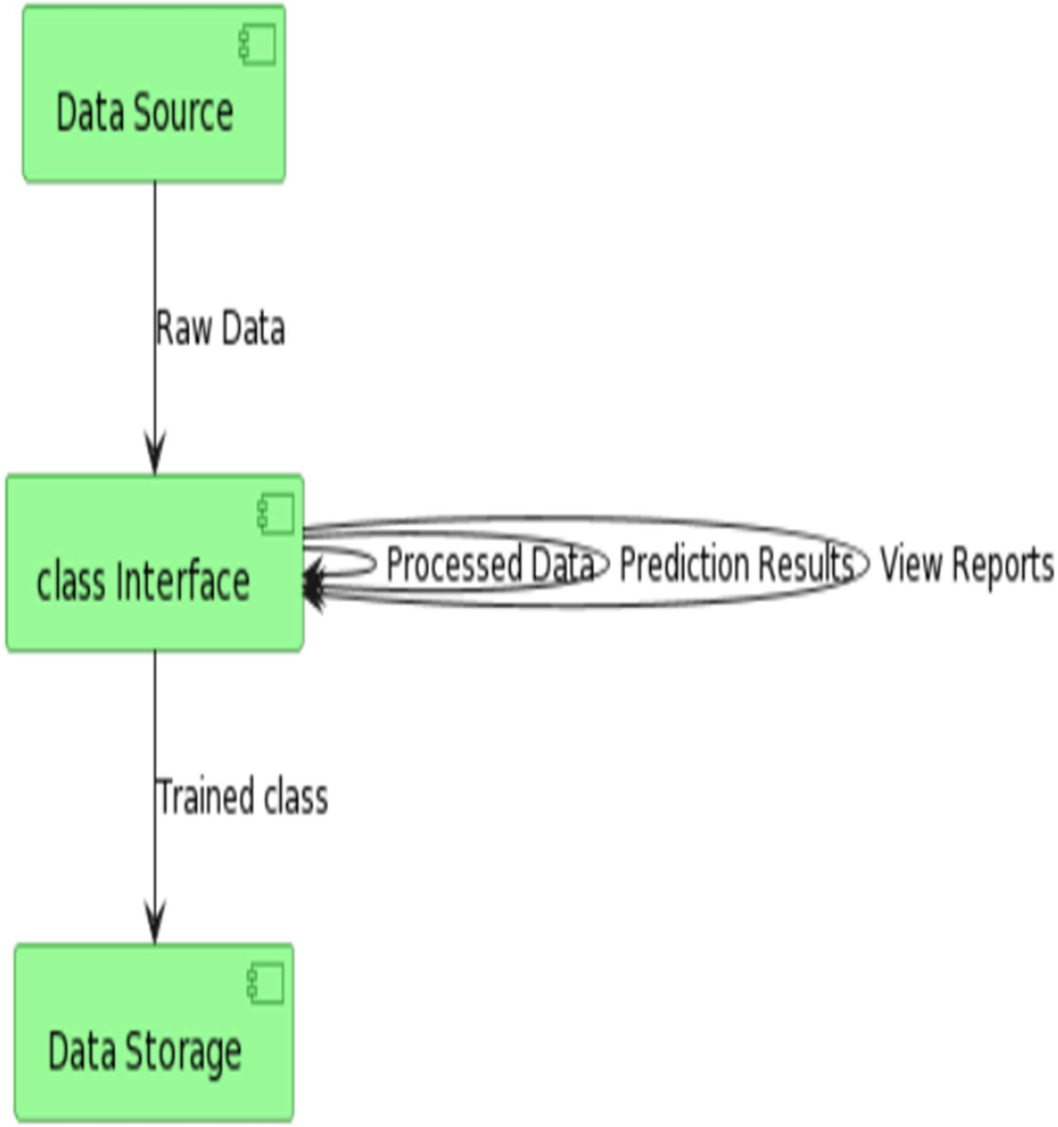
Deployment diagrams consist of nodes representing hardware devices and artifacts representing software components. Arrows indicate the deployment relationships between nodes and artifacts. Nodes could include servers, laptops, or cloud services, while artifacts might represent the database, predictive model, or user interface components.

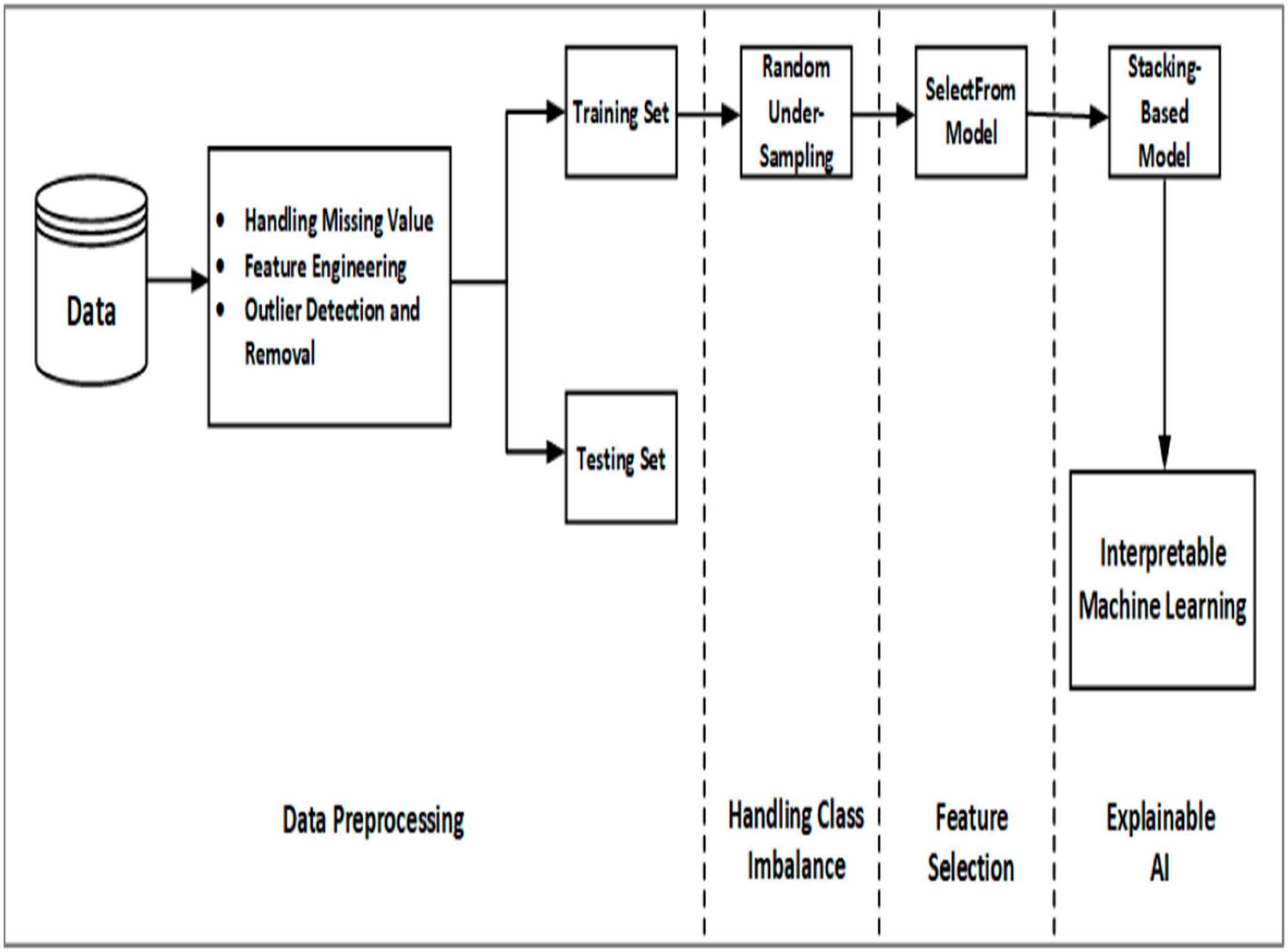
In the hospital readmission prediction project, a deployment diagram would illustrate how the developed predictive model is deployed across different nodes. For example, the model might reside on a server node, interact with a database node for data storage, and communicate with a user interface node for presenting results. The deployment diagram offers a visual guide to the physical distribution of components, aiding system architects and developers in understanding the project's infrastructure.


Prepare the datasets for readmission prediction
Apply inclusion/ exclusion criteria
Value transformation
Missing value handling
Arithmetic mean/ mode method
RF method
Standardisation/ Normalisation
Feature selection
Chi-square test
Results from previous literature --- readmission related factors
Dataset splitting
Training set
Testing set
Data balancing
Sub-training set
Validation set
Train and tune the ML model
Validate the model's prediction performance


Hospital Readmission Prediction
C class
hospitalID: int
name: string
location: string
patients: Patient[]
patientID: int
name: string
age: int
gender: string
medicalHistory: string[]
previousHospitalizations: int
readmitted: boolean
patient: Patient
probability: double
predictReadmission(): Prediction
getProbability(): double
1
1
has
generates
0..*
1


Patient
patientID: int
age: int
gender: string
medicalHistory: string
1
has
0..*
HospitalAdmission
admissionID: int
admissionDate: date
dischargeDate: date
lengthOfStay: int
dischargeStatus: string
PredictiveModel
modelID: int
trainingData: DataSet
testingData: DataSet
trainModel(): void
predictReadmission(patient: Patient): boolean
1
uses
1
DataSet
data: List<DataPoint>
addDataPoint(dataPoint: DataPoint): void
1
contains
0..*
DataPoint
predictorVariables: Map<string, Object>
outcomeVariable: boolean


Data Source
Raw Data
class Interface
Processed Data
Prediction Results
View Reports
Trained class
Data Storage



Data
Handling Missing Value
Feature Engineering
Outlier Detection and Removal
Training Set
Testing Set
Random Under-Sampling
SelectFrom Model
Stacking-Based Model
Interpretable Machine Learning
Data Preprocessing
Handling Class Imbalance
Feature Selection
Explainable AI

OUTPUTS

	0	1	2	3	4	5	6	7	8	9
encounter_id	2278392	149190	64410	500364	16680	35754	55842	63768	12522	15738
patient_nbr	8222157	55629189	86047875	82442376	42519267	82637451	84259809	114882984	48330783	63555939
race	Caucasian	Caucasian	AfricanAmerican	Caucasian	Caucasian	Caucasian	Caucasian	Caucasian	Caucasian	Caucasian
gender	Female	Female	Female	Male	Male	Male	Male	Male	Female	Female
age	[0-10)	[10-20)	[20-30)	[30-40)	[40-50)	[50-60)	[60-70)	[70-80)	[80-90)	[90-100)
weight	?	?	?	?	?	?	?	?	?	?
admission_type_id	6	1	1	1	1	2	3	1	2	3
discharge_disposition_id	25	1	1	1	1	1	1	1	1	3
admission_source_id	1	7	7	7	7	2	2	7	4	4
time_in_hospital	1	3	2	2	1	3	4	5	13	12
payer_code	?	?	?	?	?	?	?	?	?	?
medical_specialty	Pediatrics-Endocrinology	?	?	?	?	?	?	?	?	InternalMedicine
num_lab_procedures	41	59	11	44	51	31	70	73	68	33
num_procedures	0	0	5	1	0	6	1	0	2	3
num_medications	1	18	13	16	8	16	21	12	28	18
number_outpatient	0	0	2	0	0	0	0	0	0	0
number_emergency	0	0	0	0	0	0	0	0	0	0

fig-1

This graphical user interface represents an inaugural iteration created using the Tkinter library in Python. Within this interface, prominently featured are two primary functional elements, denoted as "Login" and "Register" buttons, both integrated beneath a title. It is essential to note that, in the event of a user's inaugural interaction with this interface, a prerequisite for registration is imperative to gain access and utilize the application's functionality effectively.

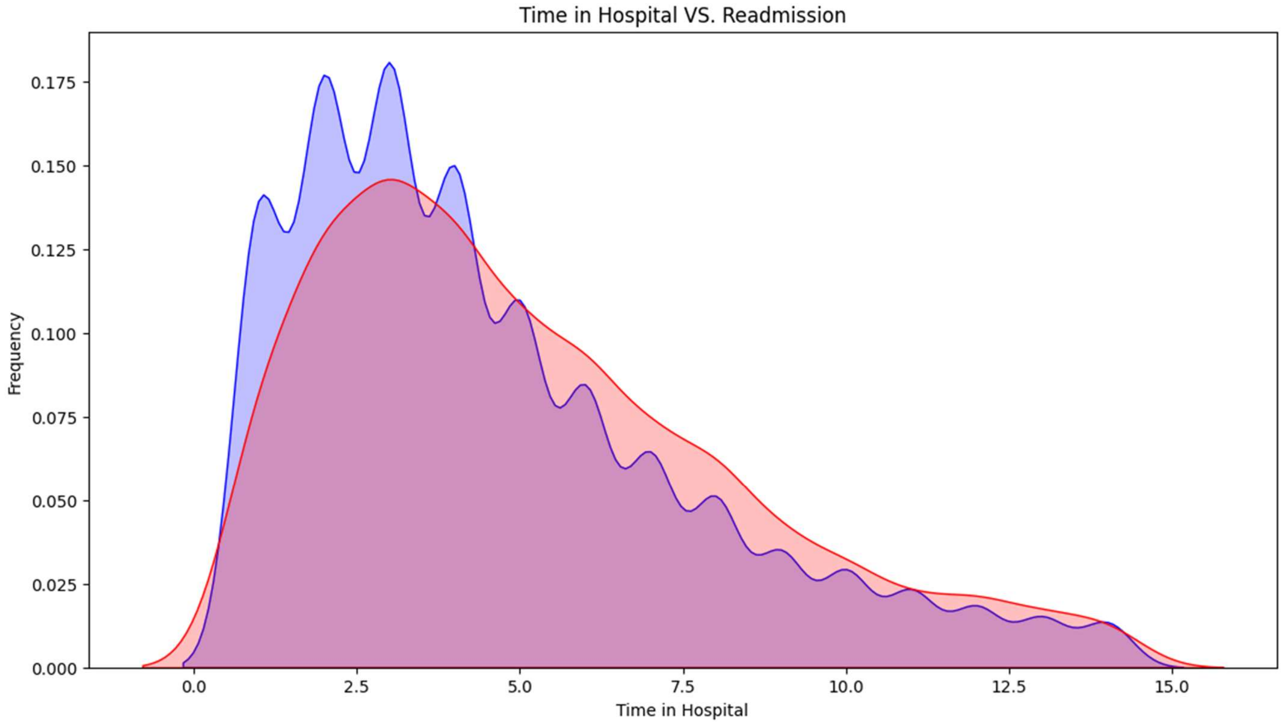


fig-2

Upon selecting the "Register" button, a new dialog window will be triggered, featuring a set of input fields, namely: "Username," "Password," "Email address," "Favorite animal," "Preferred color," and the "User's most cherished possession" as per their preference. Following the input of pertinent information into these designated fields, the registration process can be finalized by either pressing the "Enter" key or an equivalent action, thereby completing the user registration process.

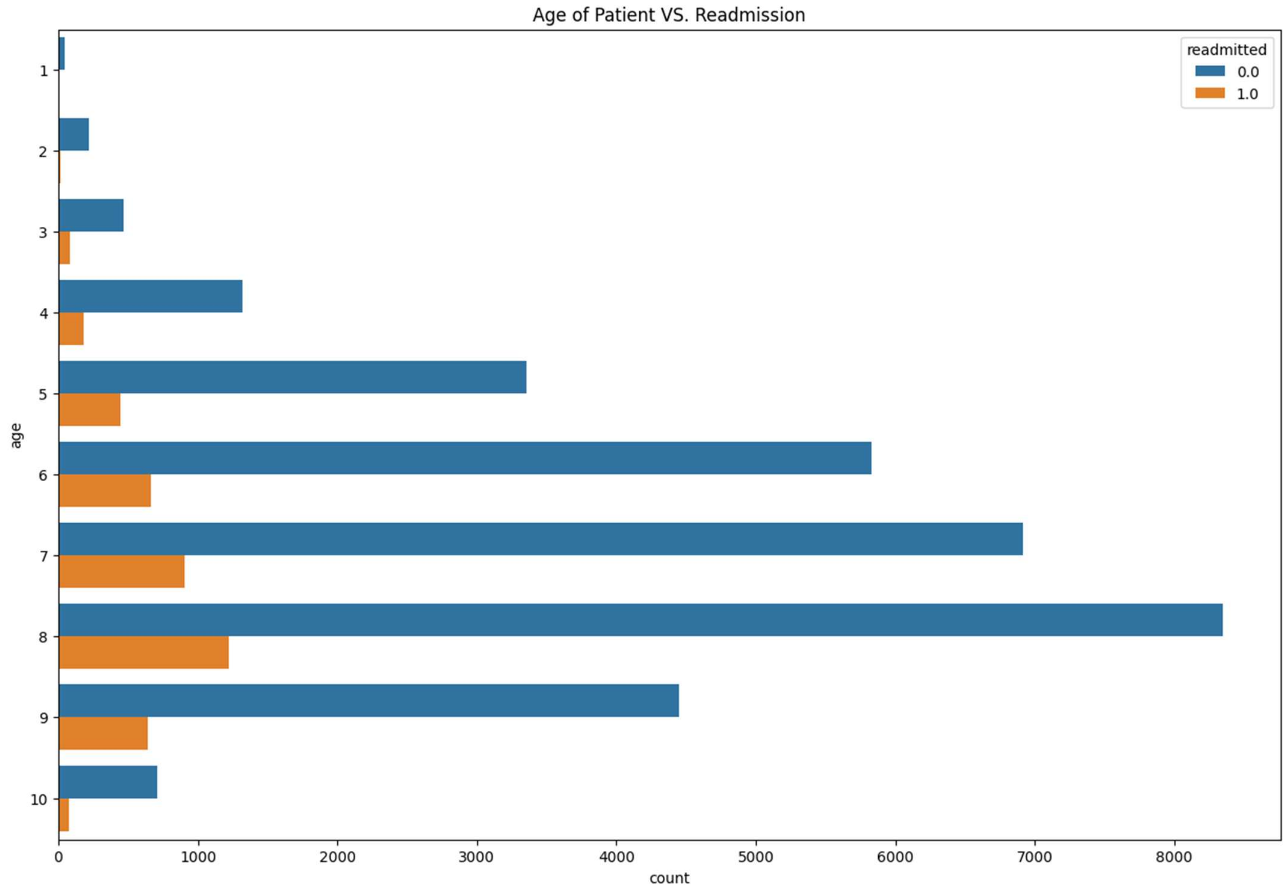


fig-3

This image serves as an exemplar illustrating the proper interrelation and completion of the respective fields. It demonstrates the ideal configuration and fulfillment of the associated elements. The visual representation provided within this image elucidates the prescribed manner in which these fields ought to be interconnected and populated. It serves as a visual reference point, elucidating the recommended standards for the harmonious association and comprehensive occupation of the designated fields. In essence, this image functions as a guiding template, conveying the preferred practices for ensuring a seamless and thorough connection between these domains, as well as the meticulous completion of each field in accordance with established guidelines and expectations.

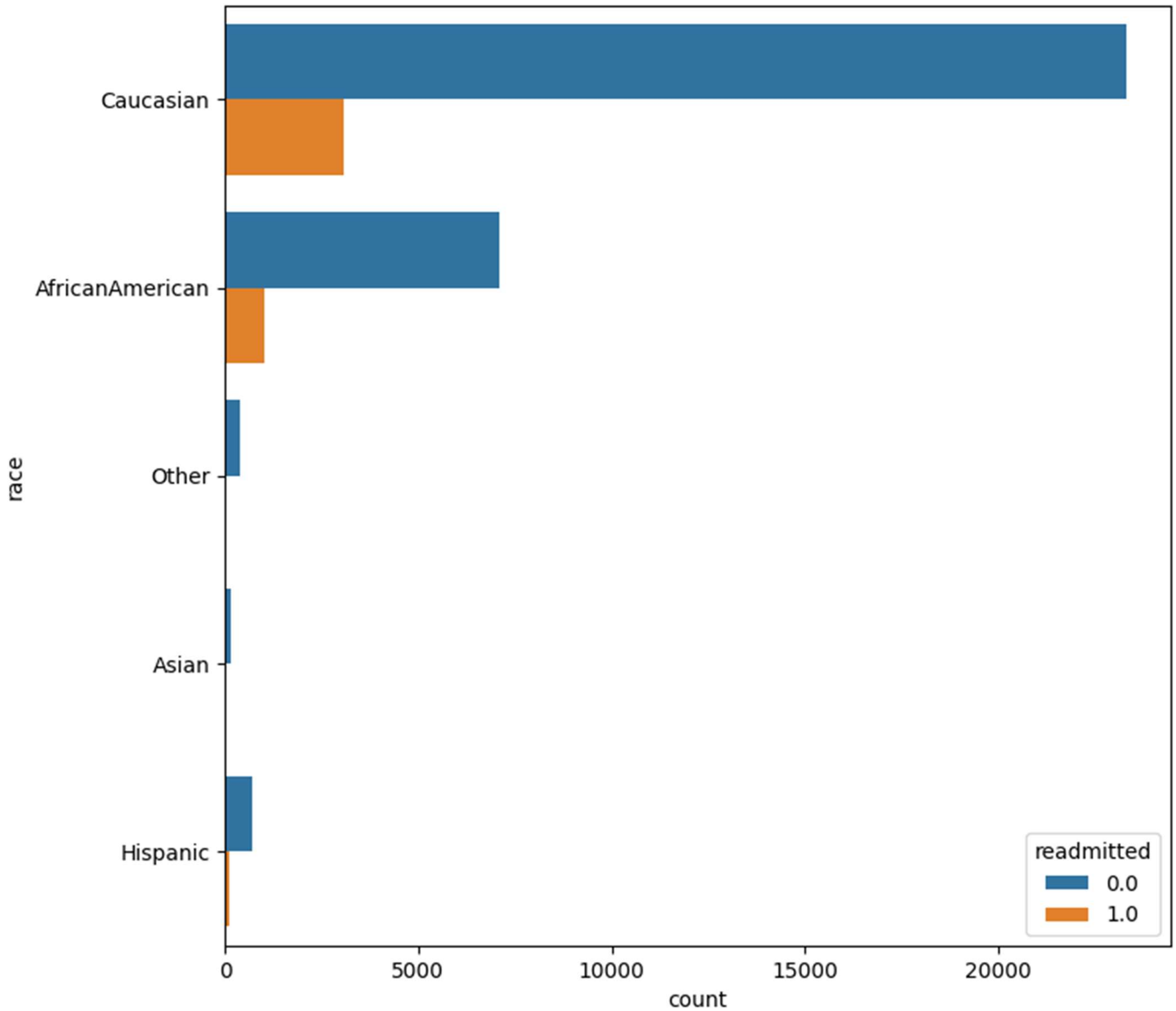


fig-4

Upon completion of the data entry process, it is recommended to finalize the operation by pressing the "Enter" key. Subsequently, a diminutive dialogue box shall manifest on the user interface, conveying a message indicating the successful addition of the data. This message serves as a confirmation of the successful execution of the data input procedure, thereby ensuring that the user is duly informed of the operation's completion. This professional and succinct notification assists in maintaining a streamlined and efficient workflow, as it eliminates any ambiguity and provides a clear acknowledgment of the data addition process, fostering a more user-friendly and comprehensible interaction with the software or system in question.

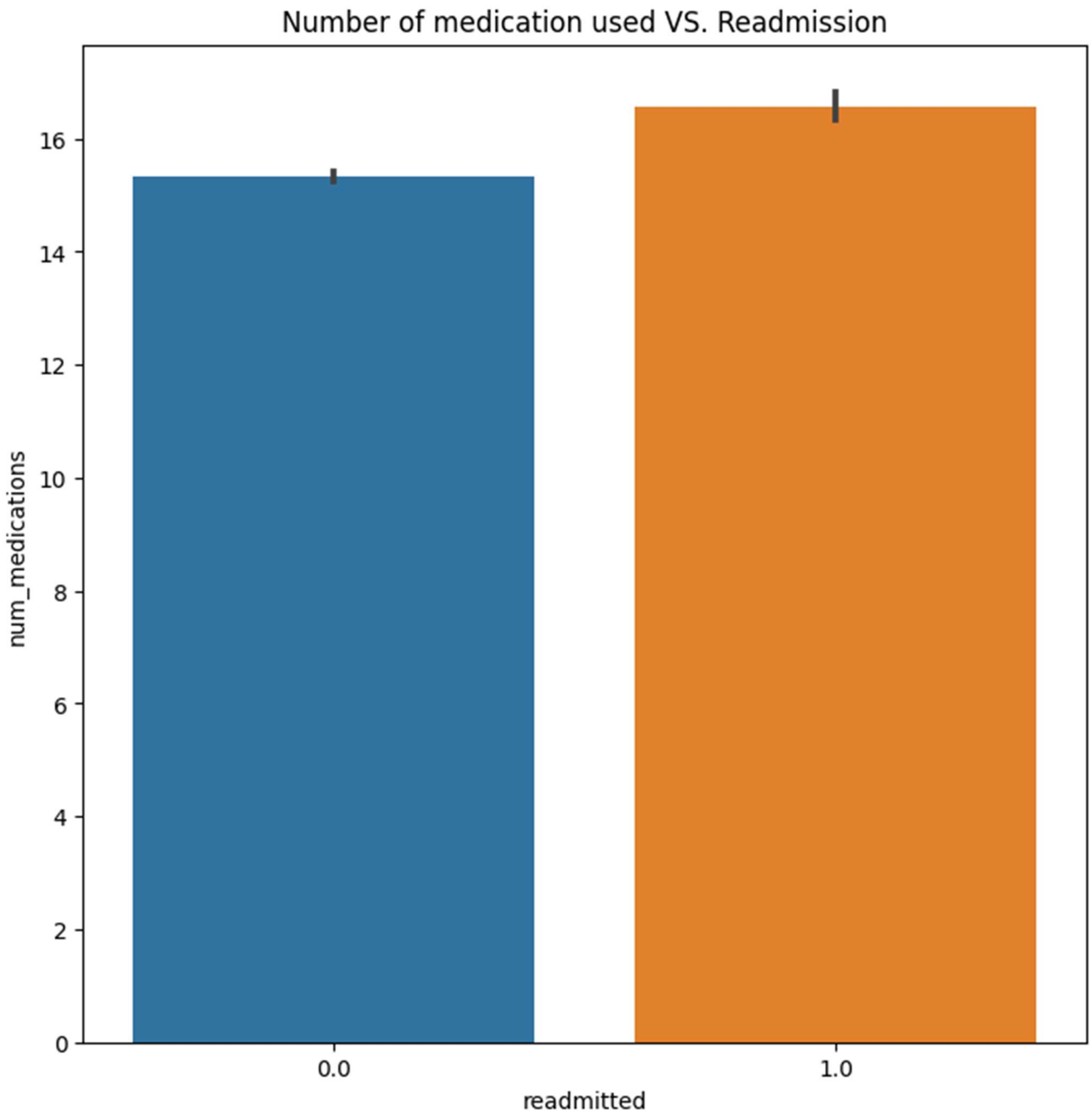


fig-5

The login interface comprises two primary input fields, specifically designated for the entry of a registered user's credentials. These fields are designed to collect essential information for user authentication, thereby facilitating access to the associated account. More specifically, the first input field is intended for the input of the user's username, a unique alphanumeric identifier that distinguishes each user within the system. The second input field is designated for the secure entry of the user's password, a confidential and sensitive string of characters serving as a key to validate the user's identity.

This carefully crafted design fosters a secure and efficient login process, ensuring that individuals with authorized access to the system can readily provide the necessary identification information to gain entry. These two fields, collectively forming the login window, are pivotal components in the user authentication process, contributing to the overall security and user experience of the platform.

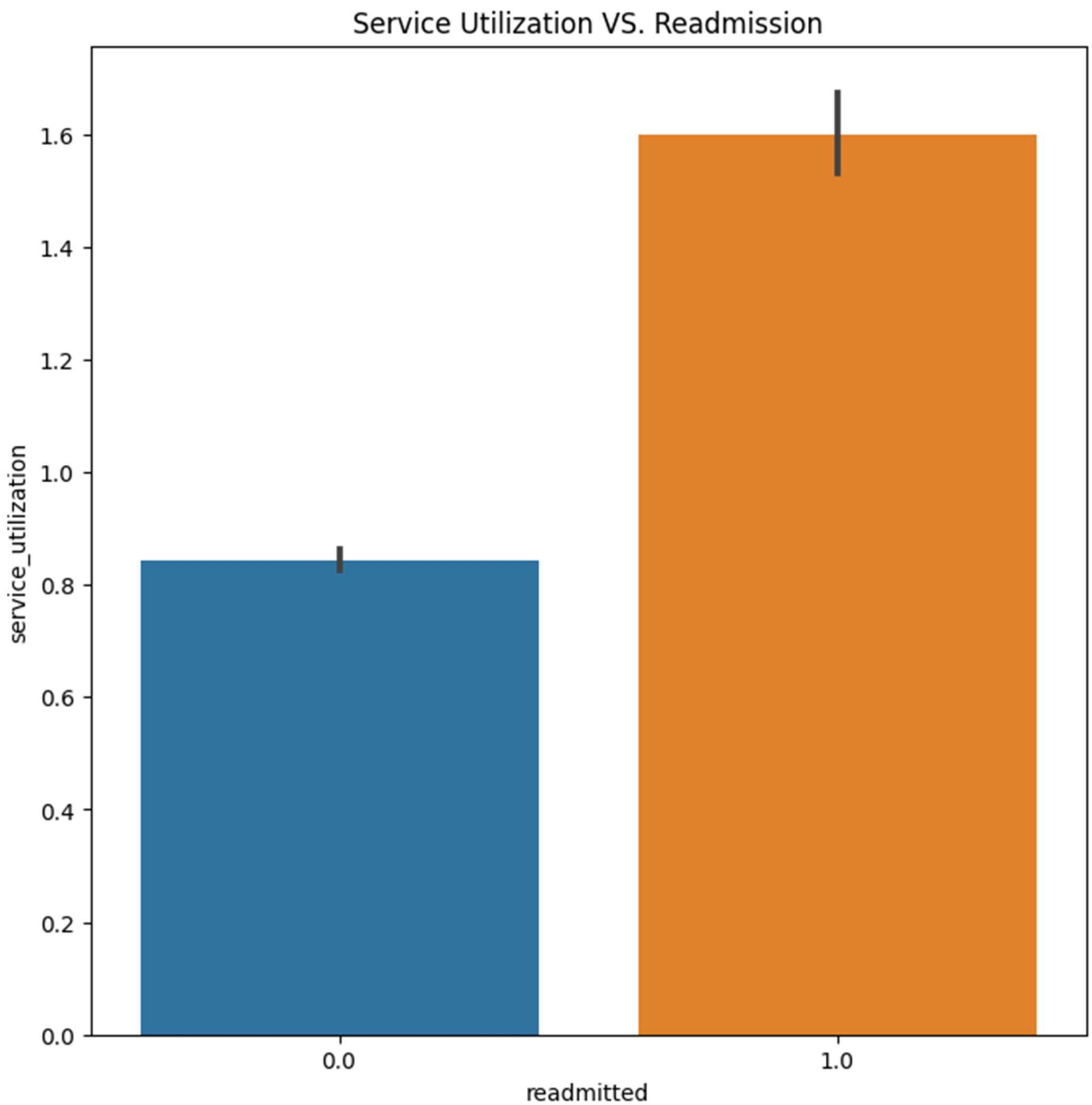


fig-6

This image serves as an illustrative depiction of the ideal method for filling out a document or form, particularly in the context of implementing initial password protection measures. It effectively demonstrates the prescribed and recommended approach for ensuring that sensitive information is safeguarded through password-based security mechanisms. This visual representation not only exemplifies the desired procedure but also conveys the fundamental importance of employing such protective measures as an initial step in securing confidential data.

In essence, this image exemplifies a model for the correct and secure execution of password protection practices, offering guidance and a visual reference point for those tasked with implementing and adhering to security protocols. Its purpose is to serve as a clear and instructive template, providing valuable insights into the foundational principles of safeguarding information through the establishment of a password-based defense mechanism, thus contributing to a heightened level of security and confidentiality in a professional context.

	age	time_in_hospital	num_lab_procedures	num_procedures	num_medications	number_diagnoses	max_glu_serum	A1Cre		
age	1.000000	0.109682	-0.027225	-0.029994	0.042562	0.209440	-0.052871	-0.072602	-0.106610	0.019594
time_in_hospital	0.109682	1.000000	0.265135	0.306499	0.508601	0.113426	0.170369	0.088769	-0.043045	-0.008691
num_lab_procedures	-0.027225	0.265135	1.000000	0.018648	-0.065680	-0.361050	0.114795	0.472975	-0.085868	0.038666
num_procedures	-0.029994	0.306499	0.018648	1.000000	0.342021	0.053421	-0.005057	-0.001858	-0.014052	-0.012593
num_medications	0.042562	0.508601	-0.065680	0.342021	1.000000	0.311148	0.181539	-0.069403	0.079108	-0.001753
number_diagnoses	0.209440	0.113426	-0.361050	0.053421	0.311148	1.000000	0.023934	-0.174221	0.009774	-0.008055
max_glu_serum	-0.052871	0.170369	0.114795	-0.005057	0.181539	0.023934	1.000000	0.050074	0.016378	-0.018013
A1Cresult	-0.072602	0.088769	0.472975	-0.001858	-0.069403	-0.174221	0.050074	1.000000	0.012519	-0.003814
metformin	-0.106610	-0.043045	-0.085868	-0.014052	0.079108	0.009774	0.016378	0.012519	1.000000	-0.008480
repaglinide	0.019594	-0.008691	0.038666	-0.012593	-0.001753	-0.008055	-0.018013	-0.003814	-0.008480	1.000000
nateglinide	0.019594	0.049887	0.061614	0.021565	0.021277	-0.018227	0.019792	-0.003814	-0.008480	-0.000357
chlorpropamide	0.014219	-0.025483	0.030777	-0.001043	-0.024250	-0.040577	0.009507	0.026783	-0.022460	-0.000944
glimepiride	0.013695	-0.003882	-0.007592	0.048745	0.027977	0.022195	-0.004793	0.001590	0.036350	-0.002491
acetohexamide	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
glipizide	0.016284	0.011559	0.007856	-0.044163	0.030288	-0.030099	0.034290	0.043261	0.064156	-0.007484
glyburide	0.105544	0.004002	-0.126645	-0.017202	0.094508	0.051691	0.022976	-0.063922	0.139388	-0.007023
tolbutamide	0.019572	-0.022578	-0.007888	-0.011956	-0.008026	0.009543	-0.009376	-0.006608	-0.014693	-0.000618

fig-7

Upon completion of the data input process, the system will initiate a validation procedure to ascertain the congruence of the provided username and password with the stored credentials. Should the system confirm a successful match, it will proceed to generate a one-time password (OTP), which will be transmitted to the previously registered email address. Subsequently, a verification step will be executed to determine whether the received email OTP corresponds to the generated OTP. In the event of a successful OTP match, the system will seamlessly transition to the subsequent operational interface or window, allowing the user to proceed with the intended tasks. This sequence of actions is designed to enhance security and ensure a streamlined user experience during the authentication and access process.

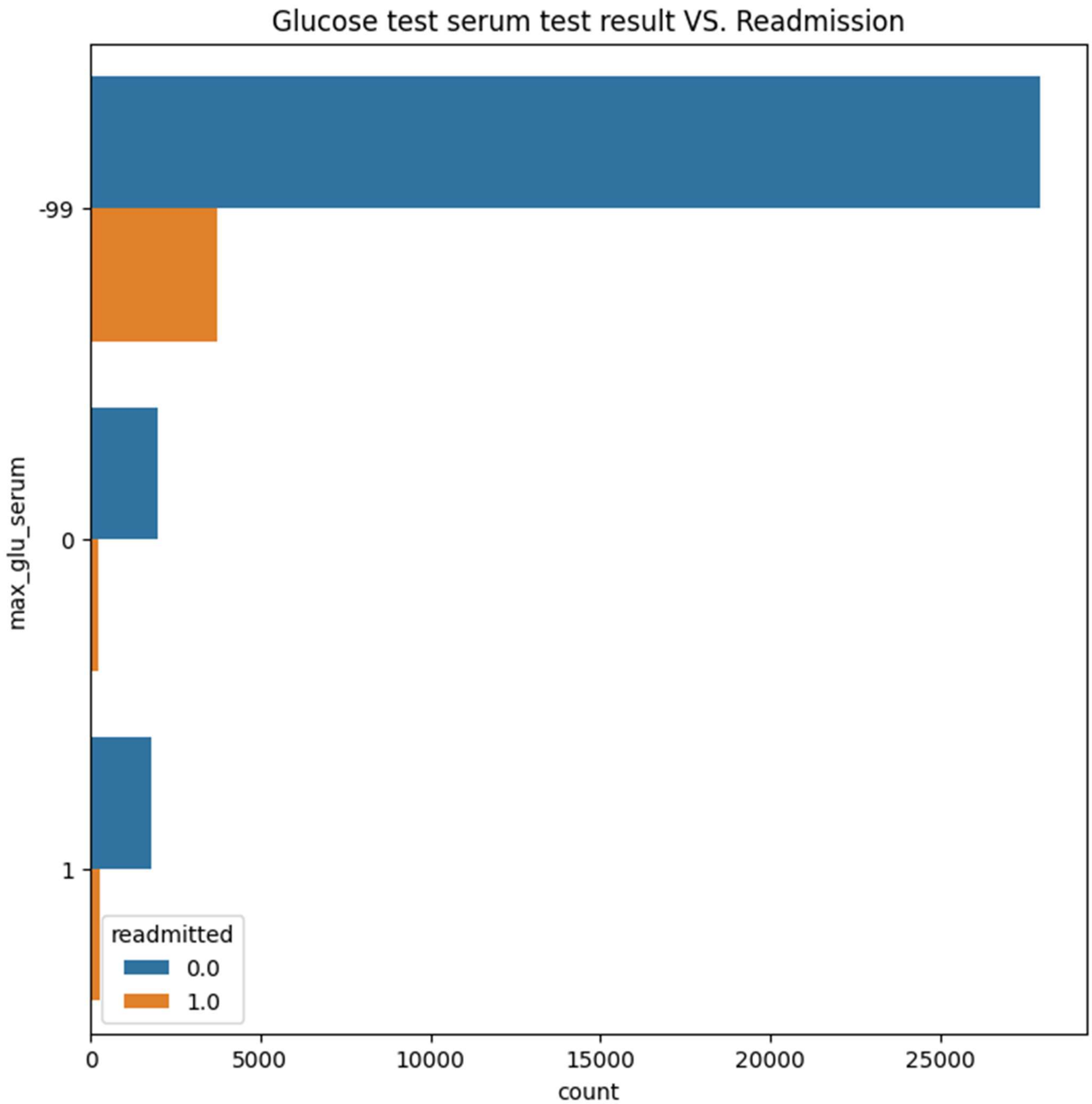


fig-8

Following successful OTP verification, the authentication process proceeds to the third step, which involves the completion of a questionnaire. Access to the system is only granted if the information provided in the questionnaire aligns with the previously provided details. This multi-step authentication method enhances security by adding an additional layer of verification, ensuring that only authorized individuals are permitted to log in.

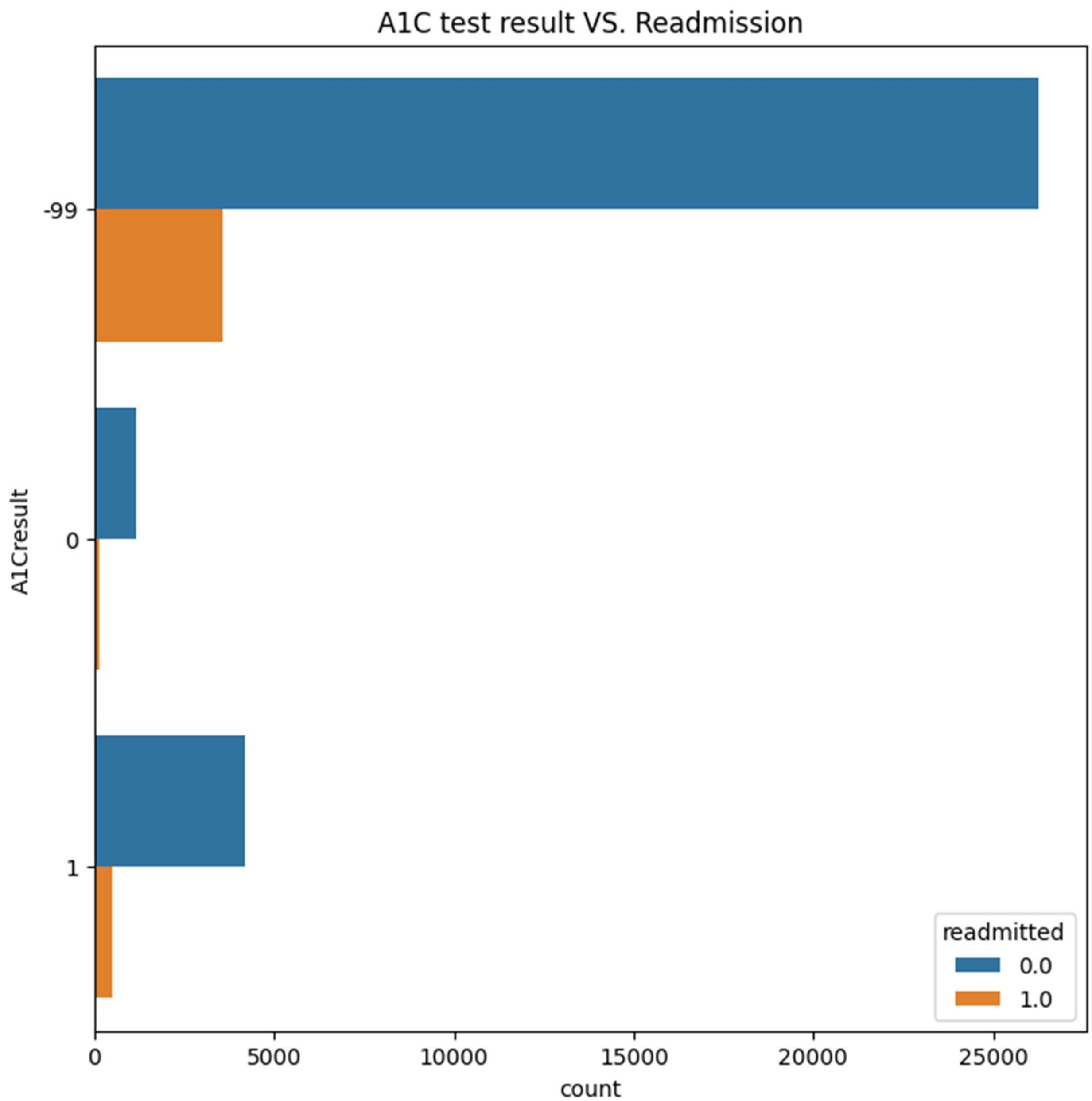
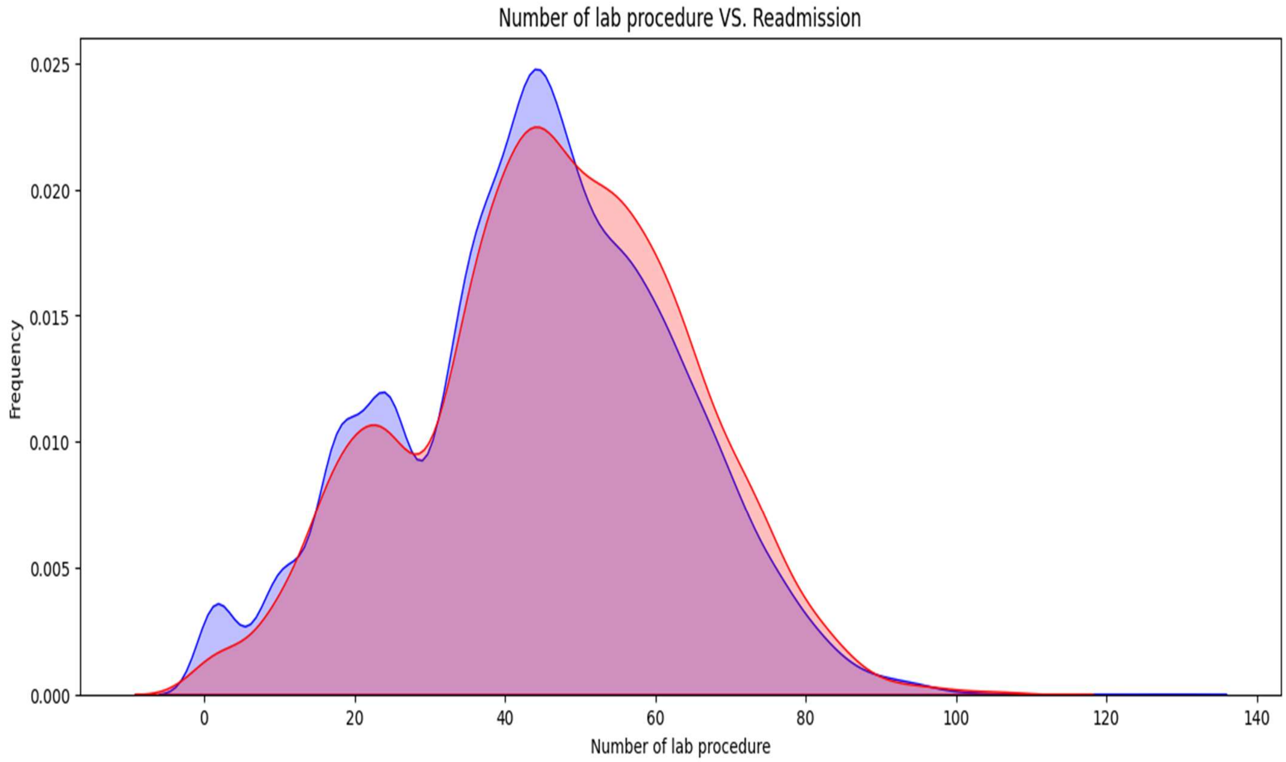


fig-9

This serves as an exemplar demonstrating the appropriate manner in which the task or form ought to be completed.



In the culmination of the authentication process, the terminal state is reached, signifying the moment at which all requisite authentication steps have been successfully concluded, thereby rendering the user prepared to initiate their lab procedure.

REFERENCES

1. Billings, J., Blunt, I., & Steventon, A. (2012). Geographical variations in hospital admissions, readmissions, and deaths in England: an analysis of linked routine data.

2. Caponnetto, P., et al. (2017). Predictors of hospital readmission of decompensated patients.

3. Chang, H. Y., et al. (2016). A novel model to predict the risk of 30-day readmission for heart failure.

4. Donzé, J., et al. (2016). Potentially avoidable 30-day hospital readmissions in medical patients.

5. Escobar, G. J., et al. (2017). Early detection of impending physiologic deterioration among patients who are not in intensive care: development of predictive models using data from an automated electronic medical record.

6. Feltner, C., et al. (2014). Transitional care interventions to prevent readmissions for persons with heart failure: a systematic review and meta-analysis.

7. Graham, K. L., et al. (2018). Risk prediction models for hospital readmission: a systematic review.

8. Hernandez, A. F., et al. (2016). Clinical research and the next generation electronic health record.

9. Joynt, K. E., & Jha, A. K. (2013). Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program.

10. Kansagara, D., et al. (2011). Risk prediction models for hospital readmission: a

systematic review.

11. Kripalani, S., et al. (2011). Predictors of readmission among elderly survivors of admission with heart failure.

12. Lavan, A. H., & Gallagher, P. (2016). Predicting risk of hospital readmission in Ireland: a statistical analysis of the Hospital Inpatient Enquiry System (HIPE).

13. Lee, J. S., et al. (2015). Identifying risk factors for 30-day readmission after hospitalization for pneumonia in the Medicare population.

14. Marcantonio, E. R., et al. (2010). Predicting 30-day mortality for pneumonia in older patients: the importance of functional status.

15. Meltzer, D. O., et al. (2010). Predicting hospital admissions and functional decline in community-dwelling older adults: the ACT aging study.

16. Morley, J. E., et al. (2011). Frailty, inflammation, and hospitalization.

17. Naylor, M. D., et al. (2011). Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial.

18. Naylor, M. D., et al. (2016). Transitional care of older adults hospitalized with heart failure: a randomized, controlled trial.

19. O'Connor, M., Hanlon, A., & Bowles, K. H. (2015). Predictors of 30-day hospital readmission following ischemic and hemorrhagic stroke.

20. Popa, C., et al. (2017). Prediction of heart failure readmission in Medicare beneficiaries: a comparison of LACE, LACE+, and the HOSPITAL score.

21. Rasmussen, S. K., et al. (2016). Predictors of 30-day readmission following ventral

hernia repair.

22. Readmissions Reduction Program (HRRP): CMS.gov.

23. Ruiz, M., et al. (2016). Predictors of early hospital readmission for decompensated heart failure.

24. Sahni, N., et al. (2015). Association of patterns of care, endovascular treatment, and outcomes in acute stroke patients with atrial fibrillation.

25. Sharma, G., et al. (2017). Predictors of early readmission among patients 40 to 64 years of age hospitalized for chronic obstructive pulmonary disease.

26. Silverstein, M. D., et al. (2011). Trends in outcomes of adult inpatients with community-acquired pneumonia.

27. Sohn, M. W., et al. (2015). Predicting all-cause risk of 30-day hospital readmission using artificial neural networks.

28. Southern, W. N., et al. (2014). Value of predictive modeling in targeting high-risk diabetic patients for intensive disease management interventions.

29. Sriram, K., et al. (2018). Predictive modeling for diabetes and hypertension outcomes.

30. Stokes, J., et al. (2017). Predictive modeling of 30-day hospital readmission rates for heart failure patients.

31. Stranges, E. A., et al. (2017). Combining predictive modeling and lean techniques for predicting 30-day readmission risk.

32. Tan, M., et al. (2019). Predicting the risk of 30-day hospital readmission among geriatric patients.

33. Tang, Y., et al. (2019). Prediction of 30-day hospital readmission for heart failure patients using a deep neural network.

34. Thorpe, C. T., et al. (2018). A computerized algorithm for predicting mortality risk and length of stay in community-acquired pneumonia.

35. van Walraven, C., et al. (2010). Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community.

36. van Walraven, C., et al. (2012). Derivation and validation of an index to predict early death or unplanned readmission after surgery.

37. van Walraven, C., et al. (2010). Derivation and validation of a clinical index for prediction of rapid HIV testing in urgent care settings.

38. van Walraven, C., et al. (2012). Derivation and validation of an index to predict early death or unplanned readmission after hospitalization for alcohol-related illnesses.

39. Varadhan, R., et al. (2013). A simple validated scale predicts readmission after acute myocardial infarction in elderly patients.

40. Wang, C., et al. (2014). A prediction model for hospital readmission of heart failure patients: a retrospective cohort study.

41. Weissman, G. E., et al. (2018). Risk prediction model for hospital readmission among older adults with serious illness.

42. Wieske, L., et al. (2015). Predictors of health-related quality of life after intensive care or intermediate care admission.

43. Williams, M. V., et al. (2012). Predictors of poor outcomes in patients discharged from

the hospital to the care of primary care physicians.

44. Wynants, L., et al. (2018). Prediction models for diagnosis and prognosis of COVID-19 infection: systematic review and critical appraisal.

45. Wynants, L., et al. (2019). Prediction models for diagnosis and prognosis of COVID-19 infection: systematic review and

critical appraisal.

46. Yan, C., et al. (2015). Predictors of hospital readmission in patients with type 2 diabetes: a retrospective cohort study.

47. Yu, S., et al. (2019). Development and validation of a 30-day hospital readmission risk model based on electronic health record information.

48. Zhang, Y., et al. (2015). Predicting 30-day all-cause hospital readmissions.

49. Zhou, H., et al. (2016). Development and validation of a risk-standardized model for predicting 30-day readmission in acute heart failure admissions.

50. Ziaeian, B., et al. (2019). Machine learning to predict the risk of 30-day heart failure readmission or death using electronic health record data.