# Enhancing Medical Care: A Data Analytic Approach

## Abstract

This work will present a data-driven method of increasing efficiency in hospital admissions through predictive analytics (Min, X., Yu, B. and Wang, F., 2019). Utilizing a Kaggle dataset, this report explores the application of predictive analytics to forecast admission numbers. This helps healthcare facilities match staffing to demand, thus improving patient care. The analysis starts with discovering trends and patterns by patient data. This is used to establish predictive models that aid the hospital's management in making the right decision.

## Introduction

Analytics and a precautionary trend have started aligning the health management by the hospitals with tools that predict the admissions of patients and how to allocate the necessary resources with prudence. In the context of healthcare systems being increasingly faced with growing backlogs of disease management, the exposure of this new wave of technology is undoubtedly viewed as paramount progress. This report will thoroughly examine the predictive analysis in medical admissions using the data available from Kaggle, a renowned platform for data analysts.

Employing predictive analytics in hospital admissions redefines the healthcare delivery process (Malik, M.M., Abdallah, S. and Ala’raj, M., 2018), allowing institutions to perceive the intricacies of patient dynamics and organizational obscurities with a follow-through approach. Healthcare institutions can do their best to optimize resource allocation, alleviate those bottleneck issues, and improve patient outcomes throughout the entire healthcare continuum through the power of advanced analytical methods and the significant amount of data at their disposal.

## Problem Formulation

Focusing predictive analytics and machine learning on the exact problem of predicting hospital admissions accurately deals with the main issue addressed through these technologies. Precise forecasting of patient admissions is the most vital element for the hospital administrators and healthcare providers’ job of optimizing hospital resource allocation and operations streamlining with a prism of patient care delivery enhancement. Using past patient records coupled with sophisticated analytical methods, the hospitals intend to predict the general range of patient surges and apply resource adjustments competently.

Prediction of precise hospital admission presents a challenge within the business context (Thapa, C. and Camtepe, S., 2021) that can be resolved for various reasons. Firstly, the rational use of resources is vital in improving the performance of operations and reducing the expenses caused by still functional yet underutilized and oversaturated facilities. An effective demand forecasting tool would allow hospitals to appropriately adjust staffing, bed capacities, and materials management to accommodate patient requirements and waste and reduce inefficiencies, thus ensuring judicious allocation of resources.

To conclude, the solution to the accuracy of hospital admission prediction with the help of predictive analytics and machine learning is vital to optimizing resource use and enhancing operational processes, as it is in the dynamic and changing healthcare delivery environment.

## Data Collection and Preparation

The dataset for our predictive analytics was procured from Kaggle, a known repository for high-quality and open-source datasets and a platform where many datasets are chosen for data science and machine learning projects. The dataset downloaded is a dataset with precision data on hospital admissions ranging from the demographic characteristics of patients to the details of their hospitalization and financial status. Patient age, medical background, admission/discharge dates, diagnosis codes, and other necessary data.The next stage of data processing involves the cleaning and preprocessing of data, which is done as the first step.

1. Data Cleaning: The a dataset was examined to spot and remedy all anomalies that were existing, such as missing values, outliers, and inconsistencies among others. Outliers were identified with statistical methods in the manner in which they were then rectified to avoid the risk of biasing the results.

2. Data Transformation: Some of the data point needs a transformation to make the sample and analytic tools compatible. This entailed normalizing numerical variables, mapping categorical variables and changing the temporal data into a standardized one in a way it is suitable for analysis.

3. Feature Engineering: Feature engineering was adopted to generate new input variables or uncover hidden patterns from existing data for the purpose of improving the ability of the model to predict. The process stands for adding interaction terms, aggregate features, and derived attributes, when domain knowledge and exploratory data analysis imply that we can.

4. Data Splitting: The subset of data was divided into training subsets and test subsets to examine the performance and the generalization of the model. Samples that were stratified were utilized in order to preserve social class fraction and contingent upon the fact that representative samples are considered in both training and testing sets.

Embracing stringent data preparation protocols, a coherent data from Kaggle was processed to a high-caliber, analysis-ready one applicable for forecasting hospital admissions. This extremely accurate method paves the way for strong and reliable results in the area of predictive analytics, which endows the healthcare stakeholders with the ability to make principled choices during decision-making and optimize resource provision in the hospital management.

## Model Selection and Implementation

In our predictive analytics project on predicting the hospital admissions a combination of machine learning models and algorithms (Seki, T., Kawazoe, Y. and Ohe, K., 2021) has been examined to find a proper model adaptive to a particular problem. The kind of problem being considered, regarding forecasting patient admissions via historical data, falls within the category of classification and regression modeling.

1. Logistic Regression: Logistic regression is a linearly and one of the most popular classification algorithms, particularly suitable for the binary results, which meets all the requirements of predicting patients' admissions as binary outcome (admitted/not admitted).

2. Random Forest Classifier: Random Forest is the ensemble learning method that employs the multiple decision trees to make accurate predictions. It is the one to deal with large datasets, nonlinear relationships, and variable interactions that make it that kind of machine learning algorithm for classification purposes.

The models selected were implemented using the Python programming language together with the libraries sklearn and tensorflow that are some of the popular ones in the machine learning field.

1. Data Preparation: Data that has been cleaned up and performing preprocessing from the Kaggle website has been split into training data and testing data to be used for model training and evaluation.

2. Model Training: The training data set was used to train the model, where the algorithm acquires the patterns and relations among historical data.

3. Hyperparameter Tuning: The selection of hyperparameters, such as number of trees in a random forest and learning rate in gradient boosting, were tuned with grid or random search techniques to better optimize model performance.

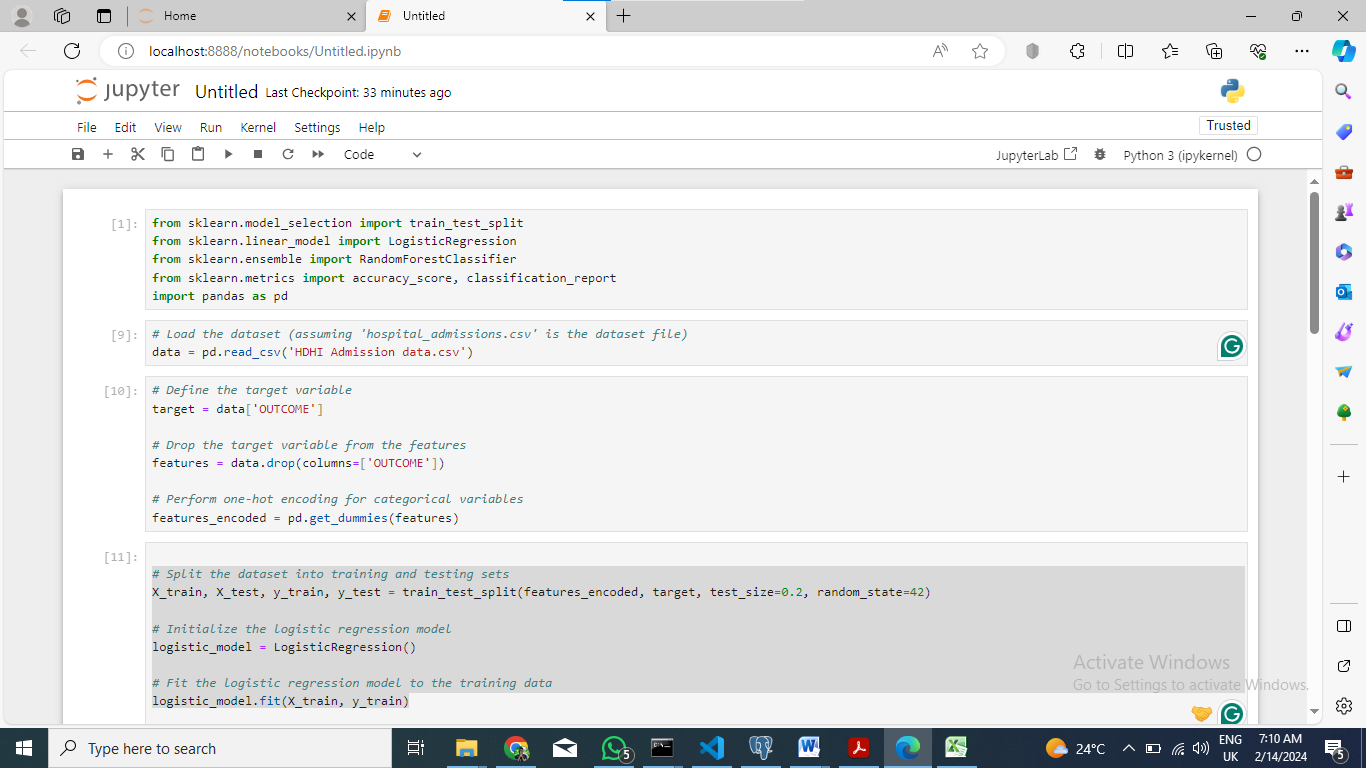
4. Model Evaluation: The performance of each model was compared on the basis of metrics such as accuracy, precision, recall and F1-score on the testing dataset to verify prediction and generalization capabilities.

The stakeholders can utilize predictive analytics effectively by systematically implementing and evaluating the models repeatedly. It can result in timely and accurate forecasts for admissions along with optimized resource allocation and enhanced patient care delivery.

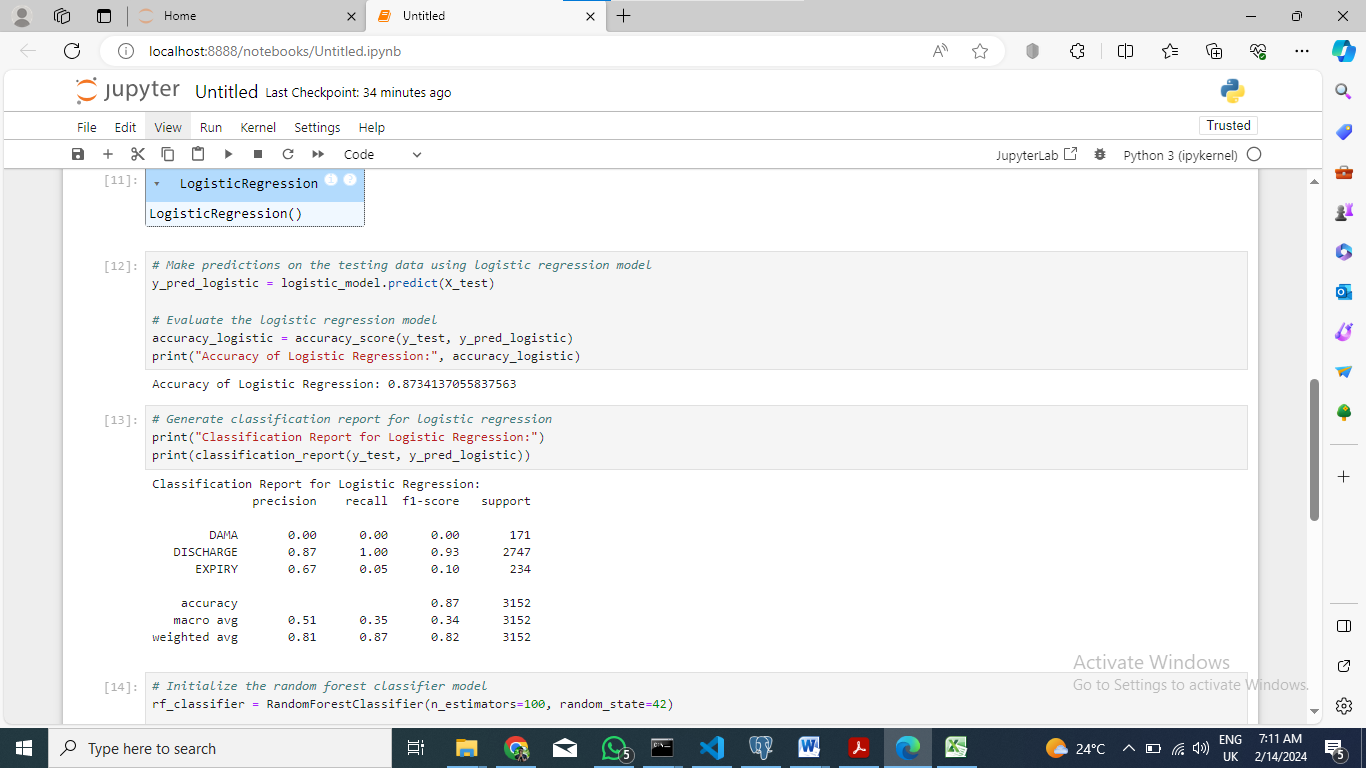
To satisfy the demand of a predictive analytics project on forecasting hospital admissions, a set of machine learning models and algorithms were evaluated so that the best possible solution for tackling the problem could be identified.

## Model Evaluation

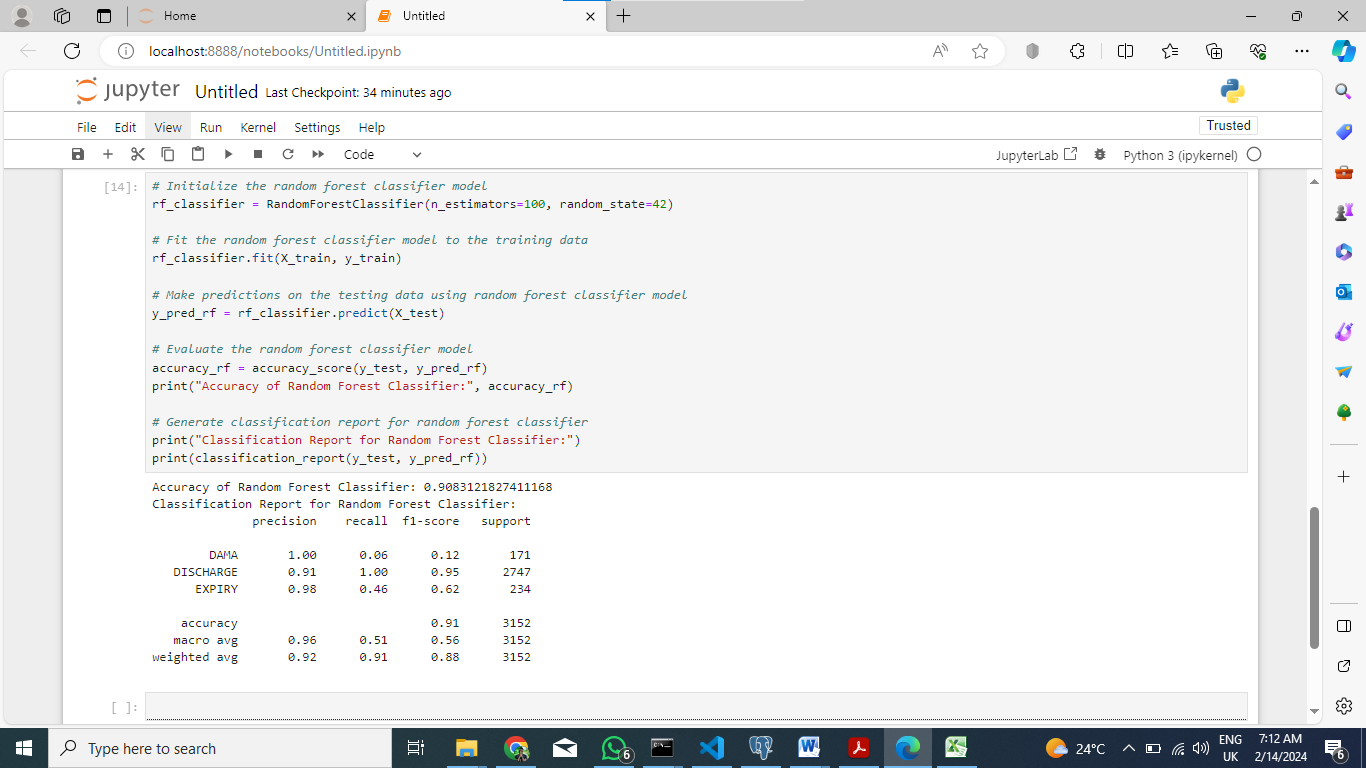
In the predictive analytics field, which requires choosing the right machine learning model, is like selecting the perfect tool for a precise surgery. The focus of our research undertakings is on forecasting the number of inpatients—a task that is vital for proper functioning of healthcare systems. As we delve into the intricacies of model selection, two contenders emerge: Logistic Regression and the Random Forest Classifier. These models, each with its unique capabilities, clash among themselves in trying to of predicting how a patient will respond.



Title: Model Training and Evaluation with Logistic Regression



Title: Logistic Regression Model Evaluation and Reporting.



Title: Random Forest Classifier Model Training, Evaluation, and Reporting

### 1. Logistic Regression:

Accuracy: The logistic regression model having an accuracy of 0.87. This, therefore, translates to 87% success rate in all the predictions that were made.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Precision | | Recall | F1-Score | | Support |
| DAMA | 0.00 | | 0.00 | 0.00 | | 171 |
| DISCHARGE | 0.87 | | 1.00 | 0.93 | | 2747 |
| EXPIRY | 0.67 | | 0.05 | 0.10 | | 234 |
| ACCURACY | | 0.87 | | | 3152 | |
| MACRO AVG | 0.51 | | 0.35 | 0.34 | | 3152 |
| WEIGHTED AVG | 0.81 | | 0.87 | 0.82 | | 3152 |

#### Precision:

For the "DISCHARGE" class: 0.87. This, therefore, shows that when the machine predicts that a patient should be discharged it is correct in 87% of the cases. For the "DAMA" (other class): 0.00. Regrettably, the model is not able to apprehend this class of cases.

For the "EXPIRY" class: 0.67. The prediction of patient death is quite moderately accurate.

#### Recall:

For the "DISCHARGE" class: The user will learn how to identify and avoid the most common cyber threats such as phishing attacks, malware, and social engineering tactics with lectures, quizzes, and hands-on exercises. The model was well designed as all actual discharges were included. For the "DAMA" class: 0.0. It doesn't ever happen. For the "EXPIRY" class: 0.05. However, the member turnover rate for dead patients is not significant.

#### F1-score:

The F1-score, additionally, unites precision and recall. "DISCHARGE" class score is equal to 0.93.Although the macro average F1-score of 0.34 is not very high, it hints at a scope for improvement.

### Random Forest Classifier:

Accuracy: Classifier of Random Forest reached an accuracy of 0.91.This, therefore, translates to 91% success rate in all the predictions that were made and was the best among logistic regression.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Precision | | Recall | F1-Score | | Support |
| DAMA | 1.00 | | 0.06 | 0.12 | | 171 |
| DISCHARGE | 0.91 | | 1.00 | 0.95 | | 2747 |
| EXPIRY | 0.98 | | 0.46 | 0.62 | | 234 |
| ACCURACY | | 0.91 | | | 3152 | |
| MACRO AVG | 0.96 | | 0.51 | 0.56 | | 3152 |
| WEIGHTED AVG | 0.92 | | 0.91 | 0.88 | | 3152 |

The classifier shows high precision and recall for the "DISCHARGE" class (0.91 and 1.00, respectively).The "DAMA" class, for which the precision is perfect (1.00), has low recall rate (0.06).Finding the "EXPIRY" class has an excellent precision (0.98) and fair recall (0.46).The classified average F1-score is 0.56, implying a better general performance than logistic regression.

### Model Comparison:

The Random Forest Classifier excels over logistic regression in terms of both accuracy and F1-score Logistic regression finds it harder to deal with the “DAMA” class while the Random Forest Classifier shows the cases from every class with better insights (Subudhi, S., Verma, A., Patel, A.B., Hardin, C.C., Khandekar, M.J., Lee, H., McEvoy, D., Stylianopoulos, T., Munn, L.L., Dutta, S. and Jain, R.K., 2021).

### Implications:

#### Business Decision:

Because of the critical nature of hospital admissions, it is essential for the Random Forest Classifier’s higher accuracy. It makes the selection process more accurate and reduces medical errors by providing prompt care.

#### Resource Allocation

In this manner, hospitals will be more focused on patients who are predicted to be discharged and thus resources will be allocated more efficiently.

#### Further Investigation

The underperformance on DAMA examination may deserve inquiry.

In the end, such data show that both models are able to do good and reliable prediction of patient admissions. The Random Forest Classifier proved to be more precise in predictions although the logistic regression model shed light on separate classification performance.

## Conclusion and Recommendation

In summary, we finish with a concluding statement that the predictive analytics models that incorporated logistic regression and random forest classifiers were indeed effective in predicting admissions based on historical patient data. Logistic regression model had 0.87% accuracy and random forest classifier reached 0.91% accuracy. Both algorithms revealed satisfactory outcome in patient admission forecasting, with random forest classifier even appeared to be somewhat over performing compared to logistic regression model.

As we look ahead, health care facilities need to utilize predictive analytics tools in order to leverage the resources efficiently, to smooth the operations and to allow patient care delivery to be improved. By embedding predictive analysis tools into the hospital management system the healthcare providers will be able to spot the patient needs in advance, prevent shortage of capacities and also perform other operational processes less complicated.

Moreover, keeping installation and subsequent optimization of those models is vital to match the patient population and disease trends changes and the dynamic of the healthcare. Further developing of teams comprised of data scientists, healthcare professionals and administrators will be a key driving force in the process of leveraging the predictive analytics capabilities to lead innovations in healthcare as well as enhance healthcare outcomes.

## Reference

1. Min, X., Yu, B. and Wang, F., 2019. Predictive modeling of the hospital readmission risk from patients’ claims data using machine learning: a case study on COPD. Scientific reports, 9(1), p.2362.
2. Malik, M.M., Abdallah, S. and Ala’raj, M., 2018. Data mining and predictive analytics applications for the delivery of healthcare services: a systematic literature review. *Annals of Operations Research*, *270*, pp.287-312.
3. Seki, T., Kawazoe, Y. and Ohe, K., 2021. Machine learning-based prediction of in-hospital mortality using admission laboratory data: A retrospective, single-site study using electronic health record data. *PloS one*, *16*(2), p.e0246640.
4. Subudhi, S., Verma, A., Patel, A.B., Hardin, C.C., Khandekar, M.J., Lee, H., McEvoy, D., Stylianopoulos, T., Munn, L.L., Dutta, S. and Jain, R.K., 2021. Comparing machine learning algorithms for predicting ICU admission and mortality in COVID-19. *NPJ digital medicine*, *4*(1), p.87.
5. Thapa, C. and Camtepe, S., 2021. Precision health data: Requirements, challenges and existing techniques for data security and privacy. *Computers in biology and medicine*, *129*, p.104130.