International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

PREDICTING PATIENT'S LENGTH OF STAY

Mr.MOHAN.K.S[1], AADISH.S[2], RATHNAVARMA.M.S.R[3],SELVAKUMAR.M[4], SWETHA.M[5]

[1] Assistant Professor, Department of Information Technology, SNS College Of Technology, Coimbatore, Tamil Nadu, India

[2, 3, 4, 5] Student, Department of Information Technology, SNS College Of Technology, Coimbatore, Tamil Nadu, India

Abstract: Healthcare organizations are under increasing pressure to improve patient care outcomes and achieve better care. While this situation represents a challenge, it also offers organizations an opportunity to dramatically improve the quality of care by leveraging more value and insights from their data. Health care analytics refer to the analysis of data using quantitative and qualitative techniques to explore trends and patterns in the acquired data. While healthcare management uses various metrics for performance, a patient's length of stay is an important one. Being able to predict the length of stay (LOS) allows hospitals to optimize their treatment plans to reduce LOS, to reduce infection rates among patients, staff, and visitors. The goal of this project is to accurately predict the Length of Stay for a patient so that the hospitals can optimize resources and function better.

Keywords: Healthcare analytics, Length of stay, Predict

1. INTRODUCTION

Extracting knowledge from databases is essential for organizations, in both private enterprises and in government agencies. If enterprises are able to recognize patterns or trends in recurrent processes, then they will be able to direct resources where they are needed, allowing a more efficient management of those available. Besides, the ability to predict events specific behaviors within some level of confidence confers additional benefits in terms of savings both at economic and human levels or managing resources.

According to a study, healthcare systems generate large amounts of administrative data about patients, departments, medical material costs, bed availability, diseases, etc. This study

departs from readily available administrative data to assess resource use in hospital systems. Concretely, a substantial amount of data stored in computer databases which, after an adequate analysis, can be helpful to improve the management of internal resources to reduce costs savings, improve patients care among other tasks. Besides, as claimed by to make a sustainable and successful integration of healthcare systems and, consequently improve not only management but the overall system, some of the main factors to consider are patients' needs, information systems data collection and performance management. Therefore, innovation should be a priority in health care, patient's care and hospital management.

A prolonged stay of patients in hospitals implies considerable costs and discomfort for patients. It also entails the need for efficient use of resources and facilities for better planning at forthcoming resources demands. These reasons motivate in depth studies that attempt to reduce the length of stay (LOS) in hospitals, as pursued elsewhere. Previous works have used hospital datasets to analyses LOS in particular departments with specific cases, and focus on the predictive effectiveness of the resulting models, but do not take into account how the context can be used to improve the models or gather additional insights. This research provides a comprehensive approach to the problem using data from all hospital departments in a large Spanish hospital located in Madrid.

The cohort consists of hospitalized patients in a period that starts on June 1st, 2010 and ends on September 29th, 2015. Here we approach the problem in global terms, analyzing all hospital departments to get an overall idea of which departments are more appropriate to assign more resources. So, predicting LOS in these departments may support more effective management and enables a more successful vision while searching for patterns about specific or generic cases in the patient's health history. In our research, machine learning techniques are applied to hospital management in an attempt 2 to optimize hospital resources more efficiently within the departments, providing an extra advantage in favor of patients and hospital entities. Thus, we report our results on predicting LOS in all departments from a hospital and present additional patterns that may complement the models with relevant insights.

2. LITERATURE SURVEY

The length of hospital stay and its implications have a significant economic and human impact. As a consequence, the prediction of that key parameter has been subject to previous research in recent years. Most previous work has analyzed length of stay in particular hospital departments within specific study groups, which has resulted in successful prediction rates, but only occasionally reporting predictive patterns. In this work we report a predictive model for length of stay (LOS) together with a study of trends and patterns that support a better understanding on how LOS varies across different hospital departments and specialties. We also analyze in which hospital departments the prediction of LOS from patient data is more insightful. After estimating predictions rates, several patterns were found; those patterns allowed, for instance, to determine how to increase prediction accuracy in women admitted to the emergency room for enteritis problems.

Overall, concerning these recognized patterns, the results are up to 21.61% better than the results with baseline machine learning algorithms in terms of error rate calculation, and up to 23.83% in terms of success rate in the number of predicted which is useful to guide the decision on where to focus attention in predicting LOS. This paper describes a smart machine learning technique to predict the patient's length of stay.

3. PROPOSED SYSTEM

In this project proposes five machine learning algorithms such us Naive bayes, Random forest classifier, XGBoost, Decision Tree and Logistic Regression. We will compare these five algorithms and Find which one is best among these algorithms. Finally, two algorithms, Such us Random Forest and Decision Tree are giving high accuracy. Random forest algorithm avoid and prevents overfitting by using multiple trees. This gives accurate and precise results. Decision trees require low computation, thus reducing time to implement and carrying low accuracy. This randomized feature selection makes random forest much more accurate than a decision tree. So we are implementing Random Forest Classifier algorithm to our project to create our model and predict the output.

4.MODULE LIST

4.1 Data collection

The data set obtained from Kaggle Website, is used in this paper for the experimental verification. In this paper we are taking some

important columns as features like "Hospital_code", "Hospital_type_code","City_Code_Hospital", "Hospital_region_code", "Available Extra Rooms in Hospital", "Department", "Ward_Type", "Ward_Facility_Code", "Bed Grade", "City_Code_Patient", "Type of Admission", "Severity of Illness", "Visitors with Patient", "Age", "Admission_Deposit" and taking "Stay" as an Output.

4.2 Data pre-processing

Pre-processing refers to the transformations applied to our data before providing the data to the algorithm. Data Preprocessing technique is used to convert the raw data into an understandable data set. In other words, whenever the information is gathered from various sources it is collected in raw format that isn't possible for the analysis.

4.3 Training data and Test data

For choosing a model we split our dataset into train and test. Here data's are split into 3:1 ratio that means Training data having 70 percent and testing data having 30 percent. In this split process preforming based on train_test_split model. After splitting we get xtrain xtest and ytrain ytest

4.4 Model creation

Contextualize machine learning in your organization. Explore the data and choose the type of algorithm. Prepare and clean the dataset. Split the prepared dataset and perform cross validation. Perform machine learning optimization. Deploy the model.

4.5 Model prediction

Predictive modelling largely overlaps with the field of machine learning. There are two types of predictive models. They are Classification models, that predict class membership, and Regression models that predict a number. These models are then made up of algorithms. In this project we are using Classification method to predict the LOS(Length of Stay)

4.6 Python

Python is a good programming language for beginners. It is a high-level language, which means a programmer can focus on what to do instead of how to do it. Writing programs in Python takes less time than in some other languages.

4.6.1 NumPy

NumPy is a Python package. It stands for Numerical Python. It is a library consisting of multidimensional array objects and a collection of routines for processing of array. Numeric, the ancestor of NumPy, was developed by Jim Hugunin. Another package Num array was also developed, having some additional functionalities. In 2005, Travis Oliphant created NumPy package by incorporating the features of Num array into Numeric package. There are many contributors to this open source project.

4.6.2 Matplotlib

It is a collection of command style functions that make matplotlib work like MATLAB. Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc. In matplotlib. pyplot various states are preserved across function calls, so that it keeps track of things like the current figure and plotting area, and the plotting functions are directed to the current axes (please note that "axes & quote here and in most places in the documentation refers to the axes part of a figure and not the strict mathematical term for more than one axis)

4.6.3 Pandas

Pandas is an open-source, BSD-licensed Python library providing high- performance, easy to-use data structures and data analysis tools for the Python programming language. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

4.6.4 Sklearn

Scikit-learn is a machine learning library for Python. It features several regression, classification and clustering algorithms including SVMs, gradient boosting, k- means, random forests and DBSCAN. It is designed to work with Python Numpy and SciPy. The scikit-learn project kicked off as a Google Summer of Code (also known as GSoC) project by David Cournapeau as scikits.learn. It gets its name from “Scikit”. Scikit-learn is used to build models and it is not recommended to use it for reading, manipulating and summarizing data as there are better frameworks available for the purpose. It is open source and released under BSD license. Install Scikit Learn Scikit assumes you have a running Python 2.7 or above platform with Numpy (1.8.2 and above) and SciPY (0.13.3 and above) packages on your device. Once we have these packages installed we can proceed with the installation.

4.6.5 Flask

Flask is an API of Python that allows us to build up web-applications. It was developed by Armin Ronacher. Flask's framework is more explicit than Django's framework and is also easier to learn because it has less base code to implement a simple web-Application. A Web-Application Framework or Web Framework is the collection of modules and libraries that helps the developer to write applications without writing the low-level codes such as protocols, thread management, etc. Flask is based on WSGI (Web Server Gateway Interface) toolkit and Jinja2 template engine.

5. DESIGN AND BLOCK DIAGRAM

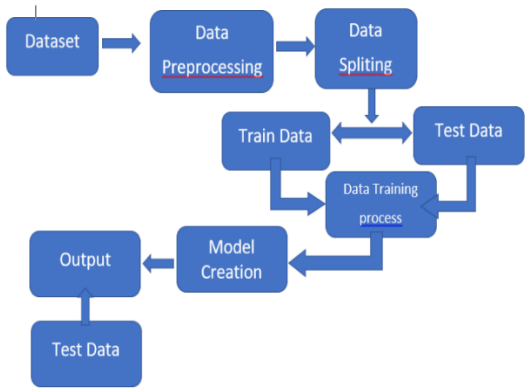

Figure 1.1 Basic machine learning diagram

5.CONCLUSION

In this project, different variables were analyzed that correlate with Length of Stay (LOS) by using patient-level and hospital-level data. By predicting a patient's length of stay at the time of admission helps hospitals to allocate resources more efficiently and manage their patients more effectively. Identifying factors that associate with LOS to predict and manage the number of days patients stay, could help hospitals in managing resources and in the development of new treatment plans. Effective use of hospital resources and reducing the length of stay can reduce overall national medical expenses. Finally we have successfully compared five algorithms and find the accuracy of these five algorithms. Decision Tree gives accuracy about 84% Naive Bayes Gives accuracy about 35%. Logistic Regression gives 36% and Random Forest Classifier Gives Same like Decision Tree algorithm which is 84%. But Random Forest Algorithm have some advantages comparing with Decision Tree. So we are using Random Forest Classifier to Train to test our model.

6. FUTURE ENHANCEMENT

The prototype can predict the length of stay using various factors which can help in resource management for financial efficiency in a dynamic health care market seeing increased regulations and costs. The resources can be blocked even before the patient starts with the treatment. The hospitals can be proactive in terms of resource management and avoid any risks when lives are at stack.

REFERENCE

1) S. Aghajani and M. Kargari, ‘‘Determining factors influencing length of stay and predicting length of stay using data mining in the general surgery department,’’Hospital Practices Res., vol. 1, no. 2, pp. 51–56, May 2016.
2) S. Barnes, E. Hamrock, M. Toerper, S. Siddiqui, and S. Levin, ‘‘Real-time prediction of inpatient length of stay for discharge prioritization,’’ J. Amer. Med. Inform. Assoc.,vol. 23, no. e1, pp. e2–e10, Apr. 2016.
3) P. Baylis, ‘‘Better health care with data mining,’’ SPSS, Shared Med. Syst. Ltd., London, U.K., 2009.
4) C.-L. Chang and P.-Y. Lu, ‘‘The study on evaluating length of hospital stay for myomectomy,’’ Int. J. Sci. Eng. Invest., vol. 5, no. 59, pp. 157–162, 2016.