CLASSIFICATION MODEL TO PREDICT DYNAMIC HEALTHCARE RESOURCE UTILIZATION AND ALLOTMENT METHOD BY USING CART ANALYSIS FOR SURGICAL PATIENTS LOS

Sandeep Ravikanti,Assistant Professor,CSE,
Mohammed Abdul Hai Siddiquie,Syed Noor ul Mustafa,Mohammad Ikramuddin,
BE VIII SEM,CSE,Methodist college of Engineering and Technology,Abids,Hyderabad,500001

Abstract: Healthcare costs and the increased demand for services especially during corona pandemic time we are facing lot of havoc which requires us to use healthcare resources and hospitalization of patients more efficiently. Static resource requirements and stay duration makes the care delivery process less efficient. We can create dynamic system by classifying patients into similar clusters by predicting stay. Developed a classification model to classify patients into various clusters of stay by using e-patients' record. There are various statistical tools for classification and prediction. However, classification and regression tree (CART) analysis is a more suitable method for analyzing healthcare data. We found that the CART analysis is also useful for determining the patient attributes that can explain the variability in resource requirements. Furthermore, we can predict the stay duration of patients based on certain factors, such as the age, the admission point, severity, emergency type and the disease type.

Keywords —ICU, LoS, Machine Learning, Data mining, CART, Bernoulli

I. INTRODUCTION

Healthcare demand is growing in several countries across the world. In General, the healthcare system comprises a mix of private and public organizations, such as hospitals, clinics, and aged care facilities. These healthcare systemsare quite affordable and accessible[1]. However, soaring healthcare costs and growing demand for services are increasing the pressure on the sustainability of the government-funded healthcare system. To be sustainable, we need to be more efficient in delivering healthcare services[2].

We can schedule the care delivery process optimally and subsequently improve the efficiency of the system if demand for services is well known. However, there is a randomness in demand for services, and it is a cause of inefficiency in the healthcare delivery process. It is possible to design a deterministic system optimally to achieve a very high, ≥ 90%, utilization of the available resources. However, in a system with intrinsic randomness[3], improving the resource utilization diminishes the quality of services. For example, if we operate an intensive care unit (ICU) at a very high, ≥ 85%, occupancy level, we may need to refuse admissions frequently because of a capacity shortage. To manage healthcare facilities efficiently, we need to minimize the effect of the randomness in demand for services on the efficiency of the system.

The random arrival time and the uncertainty in resource requirements of each individual are the sources of variability in demand for services. In hospitals, resources are bundled together, and medical professionals work in teams. A patient's resource consumption is measured by one's length of stay (LoS) at various care steps, such as the LoS in the General ward, the LoS in a surgical ward, and one's surgery duration. Therefore, the variability in resource requirements can be approximated by the variability in LoS. Moreover, in the case of elective operations, patients' arrival times are scheduled by the hospital administration. The remaining source of variability in the elective patient flow process is the randomness in LoS. We can manage a surgical suite more efficiently if we can predict patients' LoS accurately[3].

The prediction of stay duration for the distinct wards allows for effective management of the patients, the categorization[4] of patients into three types on the basis of ward type indirectly reduces the overload from the occupancy in the ward. Maintaining

a single ward for all patients is complicated, there must be a catalogue which must be maintained for all the patients to keep track of their types and the resources required as per each patient. Hence, it is optimal to maintain separate wards for different patients as per their requirement. This allows for effective understanding of the resource requirements of the patients and reduces the load as well. Furthermore, it allows us to maintain a separate catalogue for each patient. This Los is based on certain factors like age, severity, emergency type, disease type etc. It can also be used to schedule the admissions and conduct appointments.

II. LITERATURE SURVEY

The random arrival time and the uncertainty in resource requirements of each individual are the sources of variability in demand for services[1]. In hospitals[13], resources are bundled together, and medical professionals work in teams. A patient's resource consumption is measured by one's length of stay (LoS) at various care steps, such as the LoS in the general ward, the LoS in a surgical ward, and one's surgery duration[2][5]. Therefore, the variability in resource requirements can be approximated by the variability in LoS. Moreover, in the case of elective operations, patients' arrival times are scheduled by the hospital administration. The remaining source of variability in the elective patient flow process is the randomness in LoS, the variability in length of stay reduces the overall efficiency. Hence, we can manage a surgical suite more efficiently if we can predict patients' LoS accurately.

Previously there is no Model to predict the Length of Stay of patients. So, we implemented the Machine Learning Model to predict LOS of patients.Lots of resources are wasted for the Hospital Management. By using supervised Machine learning techniques, the predictions are based on the training sample containing joint observations of dependent and independent variables. Statistical techniques such as multivariate regression analysis, decision tree analysis or classification and regression tree (CART)[6] analysis, Naïve Bayes Analysis,Random Forest analysis are some of the commonly used classification techniques.This provides an advantage since predictions are made by applying the Machine learning Techniques. By predicting the LOS of patients when they admitted in Hospital, there is no chance in wastage of resources and it is very helpful for the hospital management to use the resources optimally.

2.1. System Working

The system works on the concept of supervised learning[7], the model is trained with the training dataset. The test dataset is used of accuracy analysis and testing, similarly the input of the user acts as a basis to the model for classification and regression analysis[8]. The CART algorithms predict the patient's disease and the stay duration (Los). The System Architecture can be seen in **fig 2.1**

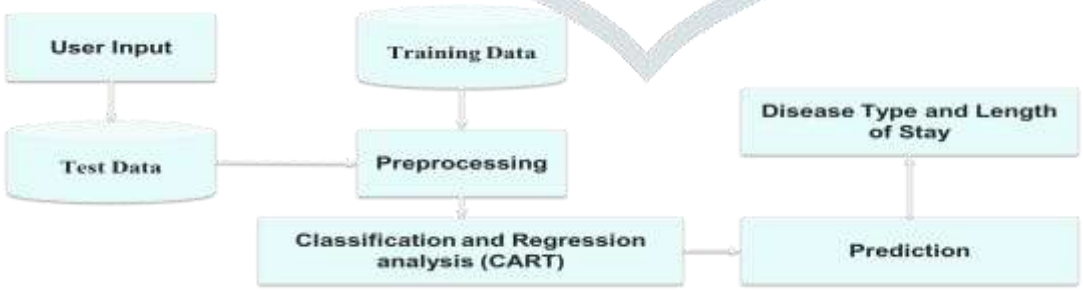

Fig 2.1 System Architecture

2.2 Algorithms

Decision Tree Algorithm

Decision Trees are a type of Supervised Machine Learning[9] where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decisions or the final outcomes and the decision nodes are where the data is split. There are two main types of Decision Trees

- Classification trees (Yes/No types)
- Regression trees (Continuous data types)

There are many algorithms out there which construct Decision Trees, but one of the best is called as ID3 Algorithm. ID3 Stands for Iterative Dichotomiser3[10].**Entropy**, also called as Shannon Entropy is denoted by H(S) for a finite set S, is the measure of the amount of uncertainty or randomness in data.

$$H(S) = \sum_{x \in X} p(x) \log_2 \frac{1}{p(x)}$$

Intuitively, it tells us about the predictability of a certain event. **Information Gain** is also called as Kullback-Leibler divergence[11] denoted by IG (S, A) for a set S is the effective change in entropy after deciding on a particular attribute A. It measures the relative change in entropy with respect to the independent variables

$$IG(S, A) = H(S) - \sum_{i=0}^{n} P(x) * H(x)$$

where IG(S, A) is the information gain by applying feature A. H(S) is the Entropy of the entire set, while the second term calculates the Entropy after applying the feature A, where P(x) is the probability of event x.

Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning[9], which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, "Random Forest[10] is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

Naïve Bayes Algorithm

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.It is mainly used in text classification that includes a high-dimensional training dataset.Naïve Bayes Classifier[8] is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.Some popular types of Naïve Bayes Algorithms are

Optimal Naive Bayes

This classifier chooses the class that has the greatest a posteriori probability of occurrence (maximum a posteriori estimation, or MAP)[7]. As follows from the name, it really is optimal but going through all possible options is rather slow and time-consuming.

Gaussian Naive Bayes

Gaussian Bayes is based on Gaussian [9], or normal distribution. It significantly speeds up the search and, under some non-strict conditions, the error is only two times higher than in Optimal Bayes (that's good!).

Multinomial Naive Bayes

It is usually applied to document [12] classification problems. It bases its decisions on discrete features (integers), for example, on the frequency of the words present in the document.

Bernoulli Naive Bayes

Bernoulli is similar to the previous type but the predictors are Boolean variables[12]. Therefore, the parameters used to predict the class variable can only have yes or no values, for example, if a word occurs in the text or not.

III. MODULES INCORPORATED DURING IMPLEMENTATION

Supervised Classification (Data Set)

Supervised learning algorithms [5] have been applied on the test data and the output obtained is compared with the actual output.

Pandas: A panda is an open source [7], BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

Numpy: NumPy is a general-purpose array-processing package[6]. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python.

Tkinter: Tkinter is the standard GUI library for Python. Python when combined with Tkinter provides a fast and easy way to create GUI applications. Tkinter[8] provides a powerful object-oriented interface to the Tk GUI toolkit.

Scikit-learn: Scikit-learn is a free machine learning library[4] for Python. It features various algorithms like support vector machine, random forests, and k-neighbors, and it also supports Python numerical and scientific libraries like NumPy and SciPy.

IV. IMPLEMENTATION AND WORKING PROCESS

4.1 Main Technologies in implementation

Python: Python is an interpreter, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development[12], as well as for use as a scripting or glue language to connect existing components together. Python's syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse.

4.2 Working Process

The Model works on the basis of CART algorithms, these algorithms are found with in the built-in module of python “sklearn”. With the incorporation of this algorithms, the dataset[14] can be identified and utilized to predict the disease type as well as the length of stay. The “Tkinter” module that is used for GUI provides a convenient interface to the user. The User can input certain factor which acts as an input to the algorithms to give a specific outcome. The Disease is selected from the most precise outcome of three different algorithms (Decision Tree, Random Forest, Naïve Bayes) The Fig 4.2.1 illustrates the algorithm’s interface.

Fig 4.2.1 Prediction of Disease and Los

Determining the Los further enables us to better manage the occupancy of the patients in the various distinct wards. The model encompasses the distinct wards[13], they include the general ward, surgical ward and the individual rooms (Single Patient). The occupancy is effectively Utilized with the patients being specific to their wards. For example, the fig 4.2.2 displays the patients admitted in the general ward, these patients have uncommon Los. The Los of is essential in managing the resource allocation for the patients.

Fig 4.2.2 General Ward Patients' Los

The Model also provides a set of operations, these operations are used to train the model further in case of variability in patients stay, this further improves the model and it becomes more accurate in predicting the stay of the patients. This prevents overload in the occupancy and the resources utilized will be optimal.

CONCLUSION AND FUTURE SCOPE

The main aim of this project is to reduce the uncertainty in patients' resource requirements and stay duration. By developing classification and prediction models, we can lower the variability in the demand with the help of an electronic database. Determining the patients Length of Stay helps us to reduce the overload on the occupancy level in the distinct wards. The CART method groups and clusters non-identical patients on the basis of ward types. Additionally, it classifies the diseases and predicts the most precise outcome the patient might be suffering from based on the analysis of the symptoms. The Prediction Model accurately predicts the stay duration of the patients by taking certain factors as input such as age, disease type, severity,

emergency type etc. In conclusion,this Los can be used to effectively manage the occupancy of the patients and utilize the resources optimally.

For the future work, the CART models could be further modified and enhanced to classify more precisely. The prediction model only predicts from an existing electronic database; hence, the model can be developed and trained to dynamically predict the stay duration and improve the overall variability and uncertainty in resource allocation. Furthermore, investigation can be done to improve the overall model and optimize more effectively.

References

[1] AIHW, "Australia's health series no. 15. Cat. no. AUS 199," Canberra: AIHW, 2015" [Online]. Available: https://www.aihw.gov.au/reports/australias-health/ australias-health-2016/contents/summary

[2] P. R. Harper, "A framework for operational modelling of hospital resources," Health Care Manag. Sci., vol. 5, no. 3, pp. 165–173, 2002.

[3] M. Faddy, N. Graves, and A. Pettitt, "Modeling length of stay in hospital and other right skewed data: Comparison of phase-type, gamma and lognormal distributions," Value Heal., vol. 12, no. 2, pp. 309–314, 2009.

[4] A. H. Marshall and S. I. McClean, "Using coxian phase-type distributions to identify patient characteristics for duration of stay in hospital," Health Care Manag. Sci., vol. 7, no. 4, pp. 285–289, 2004.

[5] G. W. Harrison and G. J. Escobar, "Length of stay and imminent discharge probability distributions from multistage models: Variation by diagnosis, severity of illness, and hospital," Health Care Manag. Sci., vol. 13, no. 3, pp. 268–279, 2010.

[6] L. Garg, S. McClean, B. Meenan, E. El-Darzi, and P. Millard, "Clustering patient length of stay using mixtures of Gaussian models and phase type distributions," in 2009 22nd IEEE Int. Symp. Comput. Med. Syst. IEEE, aug 2009, pp. 1–7. [Online]. Available: http://ieeexplore.ieee.org/document/5255245/

[7] E.-D. Elia, A. Revlin, V. Christos, G. Florin, G. Marina, and M. Peter, Intelligent Patient Management, ser. Studies in Computational Intelligence, S. McClean, P. Millard, E. El-Darzi, and C. Nugent, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, vol. 189. [Online]. Available: http://link.springer.com/10.1007/978-3-642-00179-6

[8] X. Tang, Z. Luo, and J. C. Gardiner, "Modeling hospital length of stay by Coxian phase-type regression with heterogeneity," Stat. Med., vol. 31, no. 14, pp. 1502–1516, 2012.

[9] D. E. Clark and L. M. Ryan, "Concurrent prediction of hospital mortality and length of stay from risk factors on admission," Health Serv. Res., vol. 37, no. 3, pp. 631–645, 2002.

[10] E. e. Litvak, Managing Patient Flow in Hospitals: Strategies and Solutions, 2nd ed. Illinois Joint Commission Resources, 2010

[11] M. Rouzbahman, A. Jovicic, and M. Chignell, "Can Cluster-Boosted Regression Improve Prediction: Death and Length of Stay in the ICU?" IEEE J. Biomed. Heal. Informatics, vol. 2194, no. c, pp. 851–858, 2017. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/ epic03/wrapper.htm?arnumber=7397813

[12] S. Ridley, S. Jones, A. Shahani, W. Brampton, M. Nielsen, and K. Rowan, "Classification trees. A possible method for iso-resource grouping in intensive care," Anaesthesia, vol. 53, no. 9, pp. 833–840, 1998.

[13] An IoT & AWS Based Smart Door Authentication System for Securing Hospital Maternity Wards, Sandeep Ravikanti, K.Chinmai, Volume-65 Number-1

[14] Smart Farming: A Techno Agriculture Advancement Powered by Machine Learning, Sandeep Ravikanti , Dheeraj Ganesh, IJETT– Volume 64 Number 1 – October 2018