International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 07th Jan 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=ISSUE-1

DOI: 10.48047/IJIEMR/V12/ISSUE 01/56

Title **Predicting Postoperative Hospital Stays with a Two-Stage Model Based on Electronic Patient Data**

Volume 12, Issue 1, Pages: 597-605

Paper Authors
Mrs. M.Prathiba, V.Roopasri, P.Udaya Banu, A.Aishwarya

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Predicting Postoperative Hospital Stays with a Two-Stage Model Based on Electronic Patient Data

Mrs. M.Prathiba, Assistant professor, Dept. of Information Technology, Sridevi Women's Engineering College, Hyd. madhunala.p@gmail.com

V.Roopasri, B.Tech., Dept. of Information Technology, Sridevi Women's Engineering College, Hyd.

P.Udaya Banu, B.Tech., Dept. of Information Technology, Sridevi Women's Engineering College, Hyd.

A.Aishwarya, B.Tech., Dept. of Information Technology, Sridevi Women's Engineering College, Hyd.

ABSTRACT— Due to increasing costs and demand, we must optimise healthcare use. The unpredictable nature of the demand for available resources has a bearing on the quality of care provided. An inefficient method of doing something. The level of uncertainty must be lowered. By separating patients into groups, we are able to ensure that their resource needs are addressed. Indeed, that is the situation here. In this work, we suggest a 2-classification approach using a patient record to reliably divide people into groups based on how much of a certain resource they tend to utilise. You may do just about anything with the numerous statistics Classifying patients into subgroups has been shown to decrease resource variability. This has led to the rise of CART analysis as a preferred way of sifting through large amounts of information. provides several distinct advantages over other approaches for analysing healthcare data. For instance, it can deal with the Predictors that have complex interactions. Having a parametric nature and suffering from the curse of dimensionality. CART analysis was shown to be applicable in this context as well. Reasons for the large range in resource use results, such as patient characteristics. We also discovered that a surgical surgeon and a code, in addition to the principal recommended operation for each covariate, might account for as much as 53.43 percent of the variation in patients' durations of stay (LoS). We can better manage the flow of patients, and therefore, our productivity, if we can reduce the uncertainty in estimating patients' lengths of stay.

Index Terms— Categorization of patients; Stay Prediction Model; Models for patient flow management, data mining, and tree regression;

I. INTRODUCTION

The need for healthcare is rising, both in Australia and throughout the globe.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Australia's healthcare system, which includes hospitals, clinics, and retirement homes, is a hybrid of publicly funded and privately operated facilities. Approximately 68% of the expenditures of Australia's healthcare system are borne by the government, making it both inexpensive and easily accessible. In 2015–2016, healthcare cost the Australian economy AUD 170.4 billion, or 10% of GDP [1].

The strain on the publicly financed healthcare system is expanding as a result of both increased healthcare costs and a greater demand for services. Increasing healthcare efficiency is crucial to our continued success.

Establishing healthy job competition is essential to optimising the care delivery process and maximising its efficacy. Unpredictable service demand is a major source of waste in the healthcare industry. Deterministic systems may be optimised to make as much as 90% efficient utilisation of their inputs. Increasing resource efficiency has the opposite effect in a system with intrinsic unpredictability, lowering the quality of service provided. If the ICU's occupancy rate rises beyond 85%, it may be forced to refuse treatment to those patients who really need it. Due to unpredictable demand, healthcare facilities must be efficiently handled to the fullest degree feasible.

People's arrival timings and resource needs vary widely, creating a broad range of service demand [2]. In hospitals, professionals from different fields work together to share knowledge and resources. The amount of time a patient spends in various parts of the healthcare system, such as the intensive care unit (ICU), the surgical ward, and the operating room, affects how much money and other resources must be spent on their care (LoS). Consequently, the variation in LoS may be utilised as a proxy for the ebb and flow in resource demands. Patients' arrival times for elective operations are also coordinated by the hospital. The sole consistent source of variation in the influx of elective patients is the randomness of LoS.

Accurately predicting patients' length of stay (LoS) will allow for more efficient management of a hospital's operating room. In this study, we develop a method of classifying patients according to how often they make use of a certain service. The rest of this paper is organised as follows. In Section II, we give a literature review on LoS forecasting models and patient categorization methods. Part 3 provides a more in-depth examination of the topic at hand. More information on the categorization scheme is provided in Section IV. After our careful classification of the results, we shall summarise them in Section V. Section VI offers some last thoughts and suggestions about where to take the research next.

II. RELATED WORK

The duration of a patient's stay in the hospital, like other right-skewed data, may be modelled using the phase-type, gamma, or lognormal distributions.

There has been some discussion in the literature on the possibility of fitting multi-stage models to healthcare LoS datasets. It has been discovered that hospital LoS data cannot be adequately analysed using single-stage techniques like log-normal or gamma distributions. Faddy, Graves, and Pettitt [3] found that phase type (PH) distribution are the best suited for modelling healthcare LoS data. Marshall and Mcclean [4] claim to have grouped the hospital LoS data based on the likelihood of the Markov process absorbing each phase. Gary and Gabriel [5] developed a theoretical model to estimate the probability of an impending discharge; it is based on a multi-stage model, a combination of geometric distributions. Garg et al. [6] clustered healthcare LoS data using survival tree analysis using a mix of Gaussian and PH distributions, and they found a correlation between patient characteristics and healthcare LoS. Elia et al. [7] used a generalised linear model to compare and contrast study participants by dividing them into groups according on their level of Loss. The results of a k-mean clustering technique and a two-stage procedure were also examined. All agreed that multi-stage models performed better on healthcare LoS data and that there are innate groupings of people with similar resource needs.

Indicators of risk at the time of admission may be used to forecast mortality and duration of stay.

Academics have looked at the correlation between healthcare practitioner quality and patient demographics several times. With the use of Coxian distribution fitting and partial influence calculations, Tang, Luo, and Gardiner [8] analysed LoS data from patients. The posterior mean of the intercept and coefficient of different variables were determined using a Coxian PH linear regression. Clark and Ryan [9] used the admission risk variables and patient data to develop a piecewise exponential model that predicted hospital mortality and duration of stay (LoS). Patients undergoing knee replacement surgery had their likelihood of survival (LoS) predicted using variables such as age, gender, ethnicity, socioeconomic status, and consultant. Rouzbahman, Jovicic, and Chignell [11] used k-means clustering to enhance the accuracy of regression analysis forecasts. According to Ridley et al. [12], hospital patients may be separated into groups of similar iso-resource users using classification or regression tree (CART) studies. ICU LoS was the dependent variable, whereas admission source, age, and specialisation were the independent variables. Apollo, a widely used statistical tool created by Happer [2] to categorise patients into groups with comparable resource needs, has been updated to include CART analysis. By categorising patients according to the duration of their operations, he found that age and the nature of the procedure were the most influential factors. Ting et al. [13] used time series models to estimate future outflows.

III. METHODOLOGY

In this study, we use a two-stage method called classification and regression tree (CART) to forecast a patient's postoperative hospitalisation duration. [10] It's become more and harder for hospitals to keep up with patient demand, especially given that a patient who ends up staying in the hospital longer than planned following surgery due to a miscalculation will incur additional costs and waste valuable hospital resources like beds.

The author has developed a two-stage technique to circumvent this problem by first using classifications to make predictions about the duration of a patient's hospital stay in STAGE 1, followed by clustering in STAGE 2. Both stages are driven by the CART algorithm. CART will split data into groups with shared characteristics, allowing for more accurate classification.

Since the author has collected [8] actual patients and hospital datasets and has not published this information online, we have access to an LENGTH OF STAY (LOS) database for patients.

The proposed approach will use classification, regression, and clustering techniques to improve the accuracy of LOS predictions. Since current algorithms like

Random Forest don't use any strategic tactics, their prediction accuracy is lower than that of the proposed CART algorithm.

The following components have to be built because of this project:

This section of the programme would be used to transfer the Patient Stay Dataset.

The dataset will be read into this module, and any missing values will be filled in, any non-numerical data will be encoded, and the dataset will be divided into a train and test set.

Two-Stage CART Algorithm Execution: In STAGE 1, algorithms such Random Forest, KNN, and CART will be trained on the processing training database; in STAGE 2, all three algorithms will be trained using clustering and regression [7]. Every algorithm is trained many times using cross validation to get an average error rate.

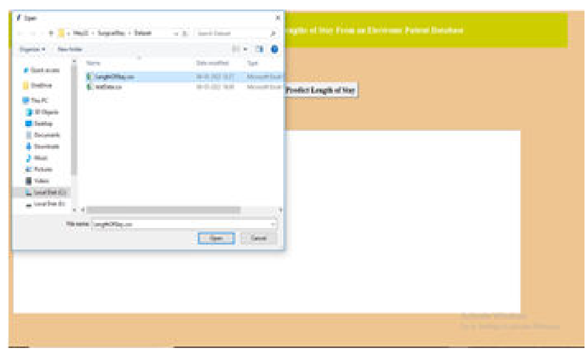
Patient Stay Prediction: After the CART algorithm has finished collecting data, it may be transferred to this module to be utilised in predicting the length of a patient's hospital stay.

Inaccurate classification rates across different algorithms may be seen in the CV Error Graph.

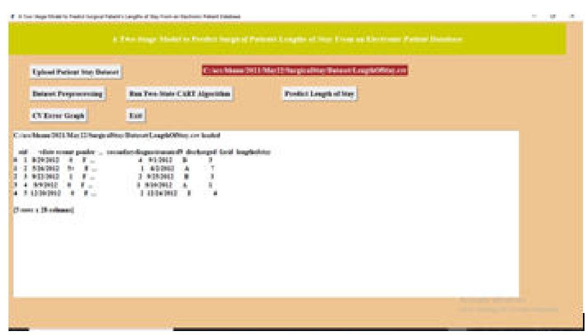
IV. RESULT AND DISCUSSION

With the goal of achieving subpar results from the project, it must be managed in order to

To upload the dataset, go to the aforementioned result and choose the tab labelled "Upload Patient Stay Dataset."



What you see below is the result of choosing a dataset from the list above, uploading it, and then clicking Open.



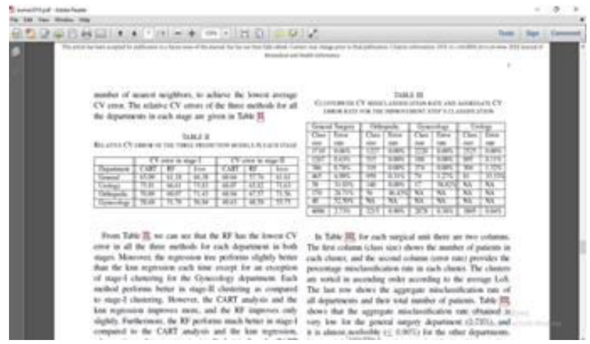
The loaded dataset had some non-numerical values, as shown in the preceding result.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

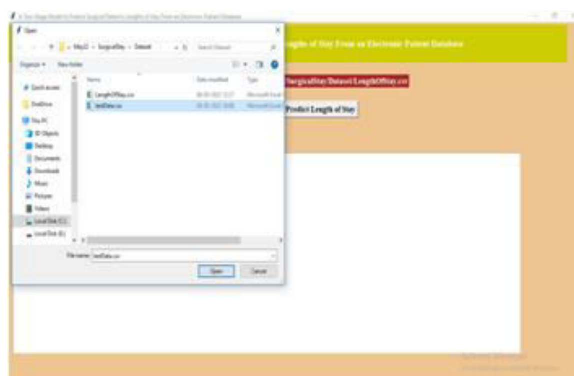
We'll have to process the dataset in order to convert or encapsulate these variables. Next, choose Dataset Pre-processing to sort, combine, and normalise the information before using it.

In the accompanying snapshot, you can see that these values have been converted to numbers. There are a total of 1000 entries in the dataset, 800 of which are utilised for training and another 200 for testing. After cleaning and organising your data, you may proceed with running the CART algorithm by selecting the Run Two-State CART Algorithm button.

On the previous screen, it is clear that Random Forest has a lower mistake rate than the suggested CART, however CART employs strategic approaches whereas Random Forest does not. As can be seen in the results below, when comparing stages 1 and 2, Random Forest has a lower error rate than the planned CART.



The error rate of random forests is shown to be less than that of the suggested CART in the first table above. Select the radio button labelled Predict Length of Stay to submit your test results.



For the aforementioned outcome, we used the Open menu item after choosing the testData.csv'file to upload and load the test

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

data.

In the result given above, the patient's test results are contained in square brackets, and the length of stay is indicated as 7 days following the arrow symbol. The expected number of days is also presented in the same format for each test record. Select the CV Error Graph button to see error plots from cross-validation.

Names of algorithms appear along the x-axis, while error rates for each of Stages 1 and 2 are shown along the y-axis. The loss rate in random forests is the lowest of the three techniques.

V. CONCLUSION

We have developed a novel method for dividing patients into more stable LoS groups by using data from electronic medical records. Using machine learning algorithms, we were able to categorise patients based on their resource requirements for strategic and operational planning purposes. CART analysis is still useful for grouping patients and carrying out feature selection even when there are many predictor variables to choose from. [6] Effective tactical and operational plans may be formulated by fitting the probability distributions generated by the CART analysis to each division.

Using the CART analysis, we were able to decrease the absolute CV inaccuracy of the stage one predictor variables by as much as 58.69 percent. Using our custom-made method, we were able to further reduce the relative CV error in forecasts by 9.0 percentage points.

The effectiveness of the CART was evaluated alongside that of the KNN and random forest regressions (RF). The results of KNN regression were similar to those of other methods, however they were less applicable. The RF projections were somewhat better than the HF's in terms of the remaining variability [11]. Thus, rather of collecting precise LoS scenario realisations, the researchers were able to derive an average value from a statistically significant subset of situations. Strategic planning and predicting new patient classes need the application of the stage-II clustering approach and the RF classification model provided in paragraph V-D, respectively.

VI. REFERENCES

[1] AIHW, "Australia's health series no. 15. Cat. no. AUS 199," Canberra: AIHW, 2015, "[Online; accessed 20-January-2018]". [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:

[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. M. Carter and H. W. Potts, "Predicting length of stay from an electronic patient record system: a primary total knee replacement example," BMC Med. Inform. Decis. Mak., vol. 14, no. 1, p. 26, dec 2014. [Online]. Available: http://bmcmedinformdecismak.biomedcentral. com/articles/10.1186/1472-6947-14-26

[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] T. Zhu, L. Luo, X. Zhang, Y. Shi, and W. Shen, "Time-Series Approaches for Forecasting the Number of Hospital Daily Discharged Inpatients," IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 2, pp. 515–526, mar 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7364163 /.