Evolution in Electrical and Electronic Engineering Vol. 3 No. 2 (2022) 687-694

EEEE

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/eeee
e-ISSN : 2756-8458

Development of An Interactive Dashboard for Outcome Prediction of a Patient Length of Stay

Muhammad Shafiy Mohamad Nasir[1], **Farhanahani Mahmud**[1,2]*

[1] Faculty of Electrical and Electronic Engineering,
Universiti Tun Hussein Onn Malaysia, 86400 Johor, MALAYSIA

[2]Microelectronics and Nanotechnology Shamsuddin Research Centre (MiNT-SRC),
Institute of Integrated Engineering (I2E),
Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, MALAYSIA

*Corresponding Author Designation

DOI: https://doi.org/10.30880/eeee.2022.03.02.081
Received 13 July 2022; Accepted 14 September 2022; Available online 30 October 2022

Abstract: Machine learning (ML) models predicting operative outcomes of mortality, and in-hospital morbidity e.g., patient length of stay and hospitalization costs can play an important role in examining the efficiency of healthcare services and resource allocation planning in a hospital. However, understanding the inner workings of the ML model in predicting the outcomes is challenging. Therefore, it is beneficial to build an interactive dashboard for interpretations of the ML model. In this study, a few regression-type supervised ML algorithms of Xtreme Gradient Boosting Regressor, Extra Trees Regressor, Random Forest Regressor, Gradient Boosting Regressor and Neural Network were analyzed to predict the patient length of stay (LOS) by using the MIMIC3d dataset. Then the finding of prediction performance was evaluated in terms of root mean squared error (RMSE) and the R-squared (R^2) performance indexes. The performance results of the ML model evaluation were compared, and the best ML model which was Xtreme Gradient Boosting Regressor with RMSE of 3.44 and R^2 of 0.992 was used to create an interactive dashboard of outcome prediction for a patient length of stay. All these works were done using Python programming. This dashboard was successfully tested in a local host, and it could be used by hospital administrators to conduct ML analysis and prediction of patient LOS patients helping them to optimize resource allocation planning in hospital units.

Keywords: Machine Learning, Xtreme Gradient Boosting Regressor, Extra Trees Regressor, Random Forest Regressor, Gradient Boosting Regressor

Corresponding author: farhanah@uthm.edu.my

1. Introduction

The worldwide proliferation of COVID-19 has increased the provision of healthcare because of the rise in the total number of patients that were admitted to the hospital. Other than the pandemic itself, there are many factors that contribute to an increase in the number of patients at the hospital. These factors include an increase in the birth rate, an uneven geographical distribution of the population, and a consistent fall in the mortality rate among births [1]-[2]. The combination of these factors results in a rise in the number of people who require medical treatment in hospitals. The hospital plays an important role in providing good quality healthcare including medical examinations. Diagnosis, treatment, and rehabilitation to the general public, particularly for patients who require long-term medical care. An operative outcome of patient length of stay (LOS) is one of the helpful indicators other than mortality and hospitalization costs which is widely measured and benchmarked as the quality indicator to improve the quality of care by hospitals, and it is also referred to in optimization of resource allocation planning in hospital units [3]-[4]. The patient’s LOS is the duration of a single episode of hospitalization. It is determined by subtracting the day of admission from the day of release to calculate the actual number of days spent in the hospital.

One of the ways that a hospital may better manage its productivity is by making use of health information technology which is considered a cost-effective strategy to boost hospital productivity by utilizing the information through analysis and processing of clinical data [5]-[6]. ML models predicting outcomes from patients’ admission for example the hospital LOS, complications, etc. can play an important role in examining the efficiency of hospital care and in optimizing hospital management better [7]-[8]. Prediction using supervised ML is preferred with better performance techniques for example random forest, artificial neural networks, and gradient boosting, although the ones that have a better performance sacrifice interpretation. While interpretable linear regression techniques would allow understanding of the reasoning behind a prediction [9]-[11]. Moreover, a dashboard that enables a view of the LOS and for analyzing the prediction is also necessary for assisting the administration team so that they can understand the inner workings of the ML model in predicting the length of stay of the patients [12]-[13].

Therefore, an interactive dashboard has been developed for the LOS prediction using a regression type of prediction to build an interactive patient LOS prediction dashboard using a supervised ML model. An aggregated MIMIC3d public dataset of Multiparameter Intelligent Monitoring in Intensive Care-III (MIMIC-III) was used to train and test the ML model to find out the potential of this approach. The web-based dashboard is an interactive dashboard for an ML model that enables the analysis and prediction of the patient LOS in a hospital.

2. Materials and Methods

2.1 Materials

This study was conducted within MIMIC3d Clinical Database. MIMIC3d dataset is an aggregated public dataset from over fifty thousand patients who stayed in critical care units at the Beth Israel Deaconess Medical Centre between 2001 and 2012 included in a freely available database of deidentified health-related data. Data on demographics, vital signs, laboratory test results, procedures, medications, caregiver notes, imaging reports, and mortality are all included in the database [14]-[15].

Meanwhile, modern data science tools, like Python and Jupyter Notebook which offer a free, time-saving learning curve as well as user-friendly data exploring, and data visualization toolkits were used in developing the ML prediction system which may help to predict the patient LOS based on MIMIC3d dataset used in this study.

2.2 Methods

Data exploration using *Pandas* profiling was used to perform exploratory data analysis efficiently. It offers a high-level API that allows a data scientist to create a detailed profile report. *Overview*, *Variables*, *Interactions*, *Correlation*, and *Missing Values* are the five key sections of the report. In a nutshell, pandas profiling saves us the time and effort of visualizing and understanding each variable's distribution. It generates a report with all of the data in one place [16].

At the data preprocessing stage, the MIMIC3d dataset was read as a data frame before the data were cleaned. The columns of input attribute with missing values; 'AdmitDiagnosis', 'religion' and 'marital_status', the columns of input attributes with document-type of data; 'AdmitDiagnosis', 'ethnicity' and 'AdmitProcedure', the column of 'ExpiredHospital' which is not relevant as one of the input attributes for the prediction, and the column of 'LOSgroupNum' were dropped from the input attributes. While the aggregated input attribute of 'TotalNumInteract' and other aggregated data of numbers was maintained as their outliers may contain meaningful information. These gave the total input attributes of 19. Since this is a regression prediction, the output label or the target is the 'LOSdays'. Next, a one-hot encoding technique was applied to the data once it was transformed from nominal to binary because the regression does not work well with nominal data. The data were then normalized, which simplifies the statistical analysis by converting many variables into a single set of values that are comparable to one another because some machine learning methods are sensitive to feature scaling and data normalization was done to improve the performance of these algorithms.

Five regression-type supervised ML models of Xtreme Gradient Boosting Regressor, Extra Trees Regressor, Random Forest Regressor, Gradient Boosting Regressor and Neural Network for the LOS prediction were analyzed in this study with 80% and 20% of the total dataset for training and testing, respectively. The analysis was conducted to select the best ML model for the interactive patient LOS prediction dashboard. The selection was according to the lowest mean squared error and root mean squared error. The five ML models were chosen for the analysis as they are among ML algorithms with a powerful approach for building supervised regression models compared to other regressor models. Figure 1 shows the workflow for the analysis of machine learning models for the LOS prediction by using Python programming code.

While Figure 2 shows the workflow of the interactive dashboard design. According to Figure 1, the development of the dashboard begins with the installation of the *ExplainerDashboard*, *Pandas*, *Dash Bootstrap* and *sklearn* libraries. The *Pandas* library is required for data manipulation and analysis using DataFrame while 80% and 20% of the total dataset and *Dash Bootstrap* libraries are for building the dashboard. To train the model for this regression problem, the *sklearn* library was used to retrieve the dataset, split it, and import the chosen regression-type ML algorithm which is XGBRegressor because it has the lowest mean squared error and root mean squared error. Next, the data were imported and split for the training is 80 % while another 20 % is for the testing. Lastly, the model was trained with the ML algorithm. Then, the dashboard was run on a local server.

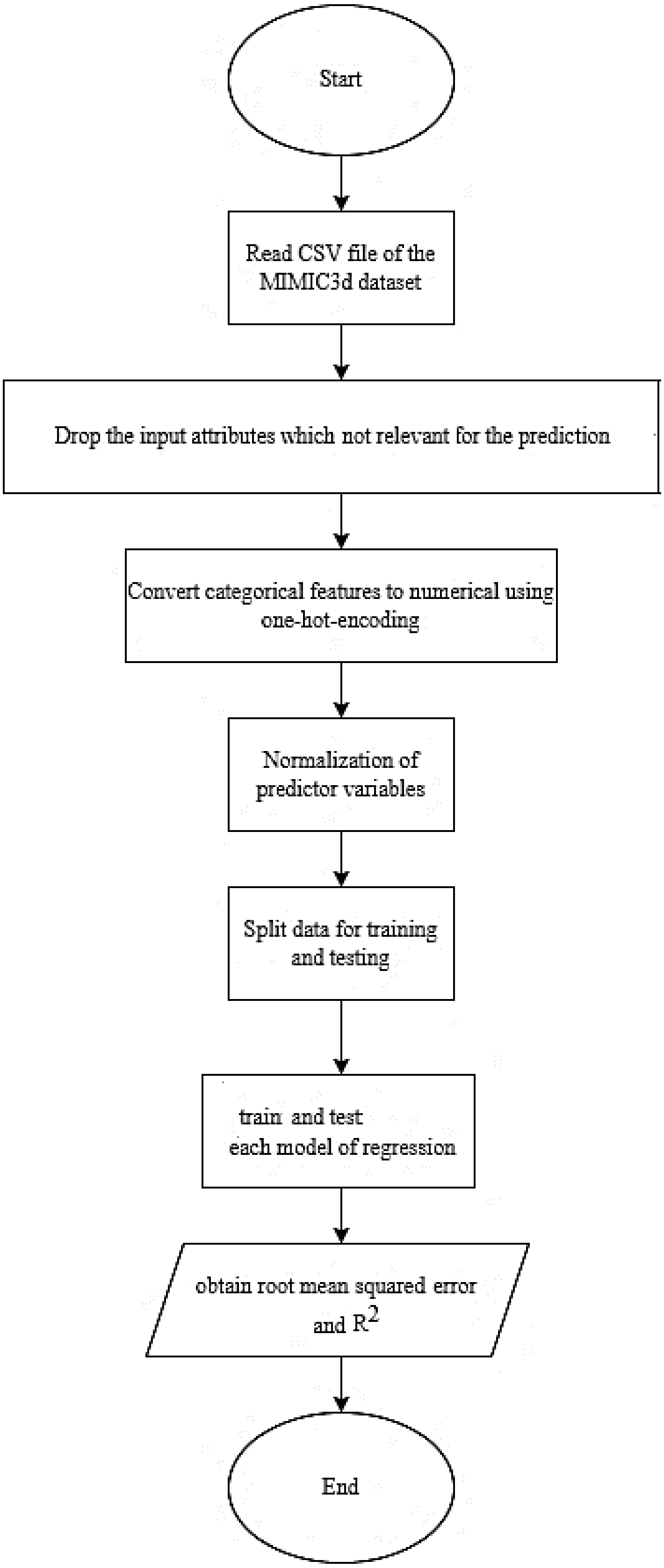

Figure 1: The workflow for the analysis of ML model for the LOS prediction by using Python

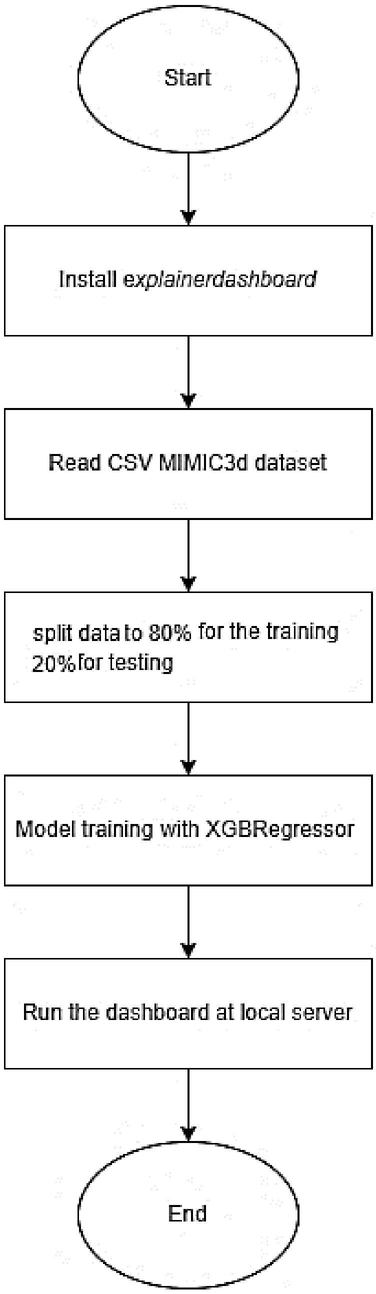

Figure 2: The workflow of the interactive dashboard design

3. Results and Discussion

In this study, the aggregated MIMIC3d from the MIMIC-III dataset was used for the machine learning modeling in predicting LOS by analyzing five regression-type machine learning algorithms which are Xtreme Gradient Boosting Regressor, Extra Trees Regressor, Random Forest Regressor, Gradient Boosting Regressor and Neural Network. Table 1 summarized the results of RMSE and the R^2 the five supervised ML algorithms. From the result of RMSE and R^2 based on the test dataset in Table 1, it is found that the Xtreme gradient Boosting Regressor gives the best performance in predicting the hospital LOS using the MIMIC3d public dataset compared to other classifiers: Extra Trees Regressor, Random Forest Regressor, Gradient Boosting Regressor and Neural Network, because it gave the lowest RMSE of 3.44 and R2, is nearest to 1 which is 0.992. Therefore, the Xtreme Gradient Boosting Regressor has been implemented into the explainer dashboard to predict the LOS. There are many advantages of Xtreme Gradient Boosting Regressor such as being highly adaptable and utilizing parallel processing power. This algorithm supports regularization and is faster than Gradient Boosting.

Table 1: The results of RMSE and the value of R^2

Item	Model	RMSE	R^2
1	Xtreme Gradient Boosting Regressor	3.44	0.99237
2	Extra Trees Regressor	3.54	1.0000
3	Random Forest Regressor	3.62	0.98077
4	Gradient Boosting Regressor	4.10	0.92082
5	Neural Network	6.95	0.73917

Then, the model was applied to the dashboard. The dashboard was developed using the *ExplainerDashboard* package in Python software that helps in conveniently and quickly building interactive dashboards for analyzing and explaining the predictions and workings of the Xtreme Gradient Boosting Regressor for the LOS prediction. Figure 3 shows a segment of the interactive dashboard interface for the LOS prediction that has been developed using *ExplainerDashboard* in Python. It shows the seven components of analysis for the user to explore in understanding the inner working of the Xtreme Gradient Boosting Regressor. The seven components are the regression performance indicators that describe how effectively the model predicts LOS day, which are *Feature Importance*, *Regression Stats*, *Individual Predictions*, *What If*, *Feature Dependence*, *Feature Interaction* and *Decision Trees*. In *What If* tab, it can show the prediction scenario by substituting the input values of features. This tab for assisting the administration team so that they can predict the length of stay of a certain patient.

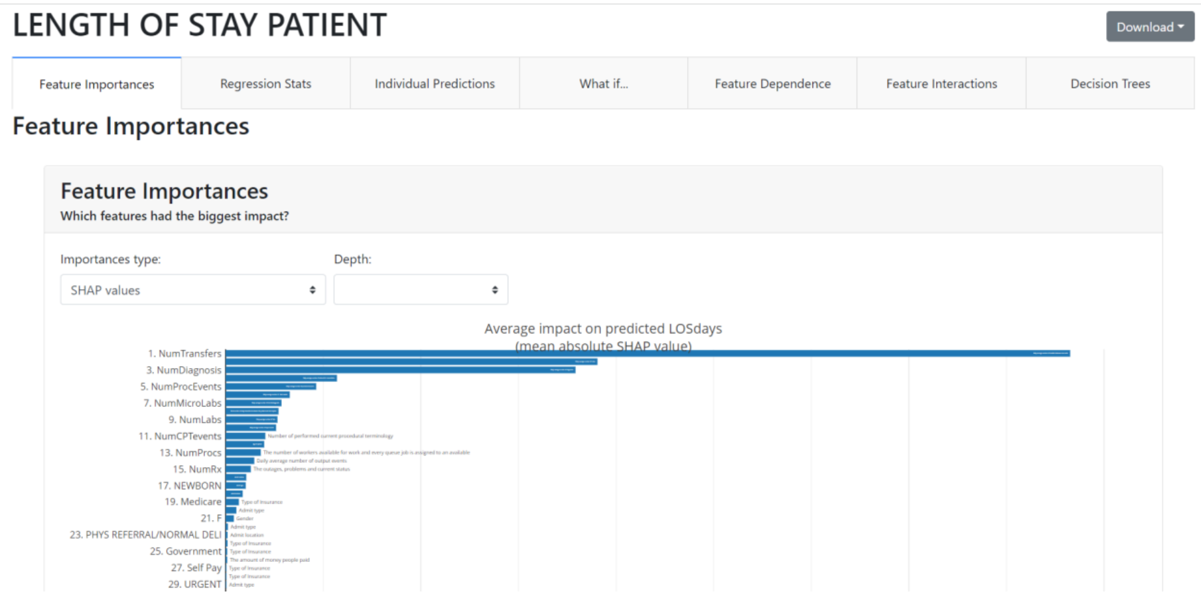

Figure 3: A segment of the Dashboard Interface

4. Conclusion

This study explored the fundamentals of machine learning and its use in predicting patient LOS. Moreover, a machine learning-based dashboard that has several interactive features were built using the *ExplainerDashboard* package in Python to enable a view of the LOS prediction and to understand the inner workings of the Xtreme Gradient Boosting Regressor machine learning model in predicting the patient LOS. Thus, this interactive prediction dashboard approach could be used to discover an ML model for predicting the operative outcomes of a patient in a certain hospital by using its recent local data to help the hospital to manage its limited resources more effectively, which will ultimately lead to an increase in their effectiveness and efficiency in providing healthcare services.

Acknowledgement

The authors would like to thank the Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia for its support in this FYP work.

References

[1] W. M. Baihaqi, M. Dianingrum, K. A. N. Ramadhan and T. Hariguna, "Linear Regression Method to Model and Forecast the Number of Patient Visits in the Hospital," 2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE), 2018, pp. 247-252, doi: 10.1109/ICITISEE.2018.8720979.

[2] R. Adawiyah, T. Badriyah and I. Syarif, “Hospital Length of Stay Prediction based on Patient Examination Using General features,” *EMITTER International Journal of Engineering Technology*, *9*(1), 2021, pp.169-181. doi: 10.24003/emitter.v9i1.609.

[3] A. Atashi, I. W. Verburg, H. Karim, M. Miri, A. Abu-Hanna, E. de Jonge, N.F. de Keizer, S. Eslami, S. “Models to predict length of stay in the Intensive Care Unit after coronary artery bypass grafting: a systematic review.” The Journal of cardiovascular surgery vol. 59(3), 2018 pp. 471-482. doi:10.23736/S0021-9509.18.09847-6.

[4] I. W. M. Verburg, A. Atashi, S. Eslami, R. Holman, A. Abu-Hanna, E. de Jonge, N. Peek, and N. F. de Keizer, “Which Models Can I Use to Predict Adult ICU Length of Stay? A Systematic Review.” Critical care medicine vol. 45(2), 2017, pp. e222-e231. doi:10.1097/CCM.0000000000002054.

[5] Y. K. Alotaibi, and F. Federico, “The impact of health information technology on patient safety.” Saudi medical journal vol. 38 (12), 2017, pp. 1173-1180. doi:10.15537/smj.2017.12.20631.

[6] M. N. Alolayyan, M. S. Alyahya, A. H. Alalawin, A. Shoukat, and F. T. Nusairat, “Health information technology and hospital performance the role of health information quality in teaching hospitals.” Heliyon vol. 6,10 e05040. 10 Oct. 2020, doi:10.1016/j.heliyon.2020.e05040.

[7] T. Exarchopoulos, E. Charitidou, P. Dedeilias, C. Charitos, and C. Routsi, “Scoring Systems for Outcome Prediction in a Cardiac Surgical Intensive Care Unit: A Comparative Study.” American journal of critical care : an official publication, American Association of Critical-Care Nurses vol. 24(4), 2015, pp. 327-34; quiz 335. doi:10.4037/ajcc2015500.

[8] S. H. Howitt, S. W. Grant, D. M. Riding, I. Malagon, C. N. McCollum, “Risk Models That Use Postoperative Patient Monitoring Data to Predict Outcomes in Adult Cardiac Surgery: A Systematic Review.” Journal of cardiothoracic and vascular anesthesia 31 (5), 2017, pp. 1865-1877.

[9] B. Panay, N. Baloian, J. A. Pino, S. Peñafiel, H. Sanson, and N. Bersano, “Predicting Health Care Costs Using Evidence Regression”, Proceedings 31, no. 1, 2019, pp. 74. https://doi.org/10.3390/proceedings2019031074.

[10] I. Eigner, F. Bodendorf, and N. Wickramasinghe, “Predicting high-cost patients by Machine Learning: A case study in an Australian private hospital group.” Proceedings of 11th

International Conference on Bioinformatics and Computational Biology, Vol. 60, 2019, pp. 94-103. doi: 10.29007/JW6H.

[11] X. Zeng, J. An, R. Lin, C. Dong, A. Zheng, J. Li, H. Duan, Q. Shu, and H. Li, “Prediction of complications after paediatric cardiac surgery.” European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery vol. 57(2), 2020, pp. 350-358. doi:10.1093/ejcts/ezz198.

[12] S. W. Marvel, J. S. House, M. Wheeler, K. Song, Y. H. Zhou, F. A. Wright, W. A. Chiu, I. Rusyn, A. Motsinger-Reif, and D. M. Reif, “The COVID-19 Pandemic Vulnerability Index (PVI) Dashboard: Monitoring County-Level Vulnerability Using Visualization, Statistical Modeling, and Machine Learning.” Environmental health perspectives vol. 129(1), 2021, 17701. doi:10.1289/EHP8690.

[13] W. H. Bagais, “Dashboard for Machine Learning Models in Health Care” (Doctoral dissertation), 2021.

[14] A. E. W. Johnson, T. J. Pollard, L. Shen L, L.W.H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovitz, L. A. Celi, R. G. Mark, “MIMIC-III, a freely accessible critical care database.” Scientific data vol. 3 160035. 24 May. 2016, doi:10.1038/sdata.2016.35.

[15] F. Mahmud, A. Z. M. Khudzari, C. P. Pau, M. F. Ramli, N. Jaffar, I. F. Gaaffar, “Pre-assessment of Machine Learning Approaches for Patient Length of Stay Prediction”, In: Mustapha, A.B., Shamsuddin, S., Zuhaib Haider Rizvi, S., Asman, S.B., Jamaian, S.S. (eds) Proceedings of the 7th International Conference on the Applications of Science and Mathematics 2021. Springer Proceedings in Physics, vol 273. Springer, Singapore. doi: 10.1007/978-981-16-8903-1_32.

[16] K. Wongsuphasawat, D. Moritz, A. Anand, J. Mackinlay, B. Howe and J. Heer, "Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations," in IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 1, pp. 649-658, 31 Jan. 2016, doi: 10.1109/TVCG.2015.2467191.