All Sciences Proceedings
All Sciences Proceedings
http://as-proceeding.com/

1st International Conference on Recent and Innovative Results in Engineering and Technology

August 16-18, 2023 : Konya, Turkey

ICRIRET
https://as-proceeding.com/index.php/icriret

Estimation of Length of Patient Stay (LOS) based on ML Algorithms

Abdulkadir Atalan*

[1]Industrial Engineering, Gaziantep Islam Science and Technology University, Gaziantep, Turkey

*(abdulkadiratalan@gmail.com)

Abstract **–**In this study, Random Forest (RF) and Adabboost (AB) algorithms from machine learning models were used to estimate the length of stay (LOS) of patients treated in a hospital. Dependent and independent variable data from 6247 patients were used for the study. The developed ML model and the AB algorithm showed the best performance. AUC, CA, F1, Prec, Recall, and MCC values for the AB model were calculated as 0.999, 0.994, 0.994, 0.994, 0.994, and 0.991, respectively. AUC, CA, F1, Prec, Recall, and MCC values of the performance measurement values of the RF algorithm were calculated as 0.982, 0.897, 0.894, 0.900, 0.897, and 0.830, respectively. ML models were developed, a patient's LOS was calculated, and bed planning for hospital management was done efficiently with the present study.

Keywords –Length of Stay, Machine Learning, Adaboost, Random Forest, Prediction

I. INTRODUCTION

One of the most important issues in hospital management, considered within the scope of health management, is resource efficiency rates [1]. Health resources are generally classified in two different ways as location and human-based [2], [3]. Employees such as doctors, nurses, civil servants, caregivers, technicians, and technicians are expressed as human-based health resources [4]. Resources such as beds, examination/treatment rooms, triage areas, and operating rooms are defined as location-based health resources [5], [6]. This study, patient stay times were analyzed by considering beds from location-based sources.

Bed occupancy rates are generally directly affected by the length of hospital stay (LOS) of the patients [7]. Bed occupancy rates are planned for the duration of a patient's stay in a hospital. For this reason, in this study, it is recommended to plan more efficiently by estimating the length of stay of the patients and calculating the occupancy rates of the beds in the hospital.

The efficiency of health resources is measured with some patient-hospital-based indicators in studies [8]–[10]. In one study, health resource numbers were evaluated by a patient's length of stay in a hospital. In another study, the number of health resources was optimized, and the patient's waiting time in the hospital was minimized [11]. Atalan et al. analyzed the number of health resources in the emergency department of a hospital with simulation models and tried to predict patient waiting times and lengths of stay [12].

Many applications such as optimization, simulation, artificial intelligence, and statistics have been used to analyze parameters such as waiting times, length of stay, and resource efficiency, among the hospital-patient indicators in the health system [13]–[15]. Generally, it aims to obtain concrete results using the discrete-event simulation method [16]–[18]. One of the most important reasons for preferring this method is that the structure of the health system or hospitals is dynamic and complex [17], [19]. However, lately, it has been observed that more than a single application is needed to solve the healthcare system's problems [20]. For this reason, the results obtained should be supported by another application. In this study, two different ML algorithms were used to test the validity of the estimation results obtained.

This study consists of four different parts. The literature part of the study is discussed in the first part. In the second part of the study, theoretical information about the methodology of this study is given. In the third part of the study, the results of the case evaluation of the research were shared. Finally, a general framework of this study was presented, and information was given about the necessity of the work.

II. Materials and Method

Describe in detail the materials and methods used when conducting the study. The citations you make from different sources must be given and referenced in references.

This study aimed to estimate the duration of patient stays by using the data of 6247 patients from a hospital. For this study, ten different inputs and one output variable are considered. The data of the study were obtained from https://github.com/duncan-wang/LOS-prediction. **Table 1** presents the definitions and properties of the variables.

Table 1. Variables used in this study and their properties

Variable	**Notation**	**Type**	**Feature 1**	**Feature 2**
Patient Gender	PtG	Input	Categoric	Discrete
age	PtA	Input	Numeric	Continuous
admit_type	AT	Input	Categoric	Discrete
admit_location	AL	Input	Categoric	Discrete
admit_diagnosis	AD	Input	Categoric	Discrete
insurance	HI	Input	Categoric	Discrete
marital_status	MS	Input	Categoric	Discrete
num_diagnosis	DN	Input	Numeric	Continuous
num_drugs	ND	Input	Numeric	Continuous
num_procedural	NP	Input	Numeric	Continuous
LOS_days	LOS	Output	Numeric	Continuous

Values, such as mean, standard deviation, variance, minimum, maximum, kurtosis, skewness, and number of samples, were calculated among the descriptive statistical criteria of dependent and independent variables. Descriptive statistics data of the variables are mentioned in **Table 2.**

Table 2. Basic statistics of variables

Variable	LOS		PtA	
PtG	F	M	F	M
Total Count	2780.0	3467.0	2780.0	3467.0
Mean	57302.0	57878.0	53654.0	53064.0
StDev	40088.0	39532.0	26544.0	25098.0
Variance	160706.0	156279.0	704570.0	629904.0
Minimum	0.0	0.0	0.0	0.0
Q1	24950.0	26152.0	43000.0	43000.0
Median	51537.0	51856.0	59000.0	58000.0
Q3	82731.0	83907.0	75000.0	72000.0
Maximum	224041.0	223676.0	88000.0	88000.0
IQR	57781.0	57755.0	32000.0	29000.0
Skewness	0.7	0.7	-0.9	-0.9
Kurtosis	0.1	-0.1	-0.2	-0.1

In this study, AB and RF models from ML algorithms were preferred to estimate the LOS values of patients. Four different performance measurement values were calculated to verify the validity of these models' estimation data. Performance measurement criteria are Area under the ROC Curve (AUC), Accuracy classification score (CA), balanced F-score or F-measure (F1), Precision (Prec), Recall, and Matthews correlation coefficient (MCC). This study used the preferred RF and AB algorithms Orange 3.34 computer program to obtain the prediction data. The ML model developed for this study is shown in **Figure 1.**

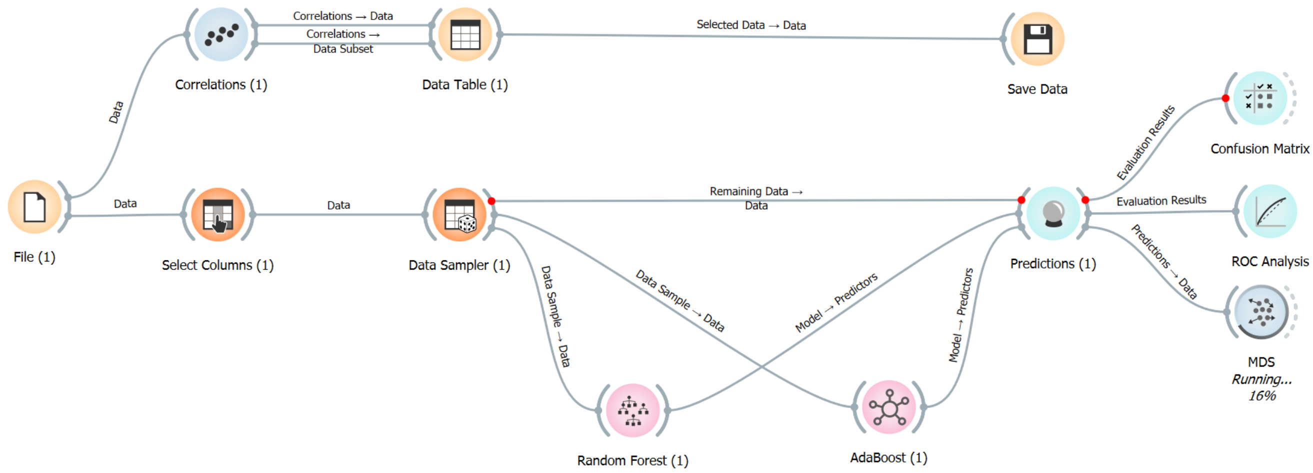

Figure 1. Screenshot of the ML model

The dependent variable LOS factor data are divided into 90% and 10% for ML algorithms' training and testing stages. The numerical results obtained by running the data used for this study in ML models are discussed in the next section.

III. RESULTS

This study aimed to estimate the data of the dependent variable by using the data of 10 independent and one dependent variable of 6425 patients. The estimation data was obtained using the RF and AB algorithms of the ML model. Performance measurement data were calculated to verify the validity of the prediction results of these algorithms. The performance measurement values of this study are given in **Table 3**.

Table 3. Performance measurement values of ML algorithms

Algorithms	RF	AB
AUC	0.982	0.999
CA	0.897	0.994
F1	0.894	0.994
Prec	0.9	0.994
Recall	0.897	0.994
MCC	0.83	0.991

According to the results obtained, the AUC, CA, F1, Prec, Recall, and MCC values of the AB model were calculated as 0.999, 0.994, 0.994, 0.994, 0.994, and 0.991, respectively. AUC, CA, F1, Prec, Recall, and MCC values of the performance measurement values of the RF algorithm were calculated as 0.982, 0.897, 0.894, 0.900, 0.897, and 0.830, respectively. According to the estimation data of these two algorithms, Confusion matrix data are given in **Tables 4** and **5.**

Table 4. Confusion matrix data according to the RF algorithm

Actual	Predicted	Long	Medium	Short	Sum of Data
	Long	95.1%	4.3%	6..0%	99
	Medium	1.7%	91.0%	6.4%	221
	Short	3.2%	4.8%	87.6%	304
Sum of Data		73	219	332	624

Tablo 5. AB algoritmasına göre confusion matrix verileri

Actual	Predicted	Long	Medium	Short	Sum of Data
	Long	99.5%	0.1%	0.3%	99
	Medium	0.1%	99.4%	0.3%	221
	Short	0.3%	0.5%	99.4%	304
Sum of Data		98	222	304	624

According to the results obtained during the testing phase of the AB and RF algorithms, the accuracy percentages of the expressions representing the duration of a patient's hospital stay were calculated as 99.4 and 87.6, respectively. According to AB and RF algorithms, the accuracy rates of a patient's long hospital stay were 99.5% and 95.1%. According to the AB and RF algorithms, the accuracy rates of a patient's mid-term hospital stay were 91.0% and 99.4%.

To obtain the prediction data, 10% was used for the testing phase. The multidimensional scaling method was used to see the distribution of the prediction data in the test phase according to LOS types. The multidimensional scaling graph of this study is shared in **Figure 2**. LOS data belonging to this scaling method are generally distributed homogeneously.

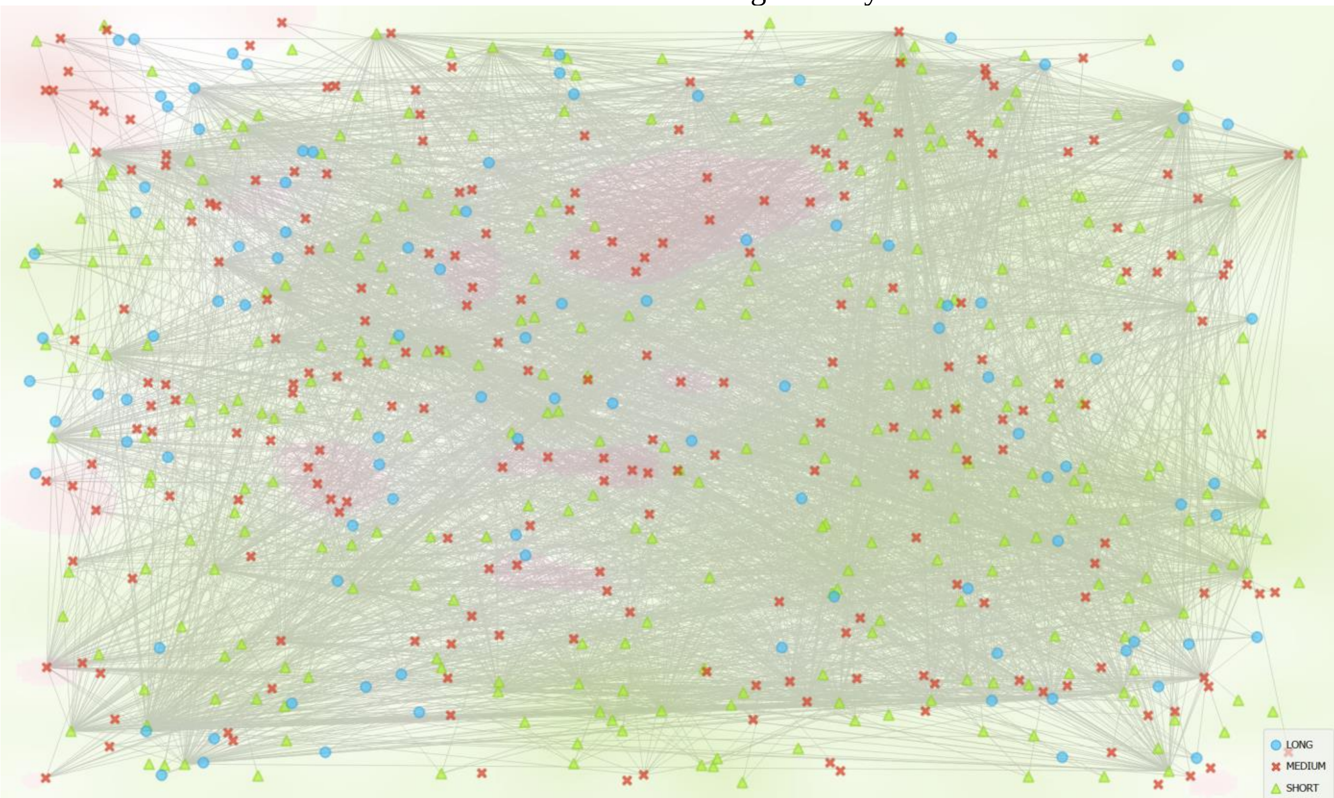

Figure 2. Multidimensional scaling plot

IV. CONCLUSION

This study ran the ML algorithms AB and RF models to estimate the LOS values of patients admitted to a hospital. To verify the validity of the estimation results obtained, the performance measurement values of the algorithms were calculated. According to these values, while the AB algorithm performs best, the RF model performs poorly despite its high accuracy. This study makes a significant contribution to the bed planning of hospital management, thanks to the estimation data to be obtained using ML algorithms. As a result of the concrete results obtained by using accurate data in this study, it was concluded that the proposed method should be used in hospital management.

REFERENCES

[1] [1] S. Ajami, S. Ketabi, M. H. Yarmohammadian, and H. Bagherian, “Waiting time in emergency department by simulation.,” Stud. Health Technol. Inform., vol. 164, pp. 196–200, 2011.

[2] [2] N. F. K. Dönmez, A. Atalan, and C. Ç. Dönmez, “Desirability Optimization Models to Create the Global Healthcare Competitiveness Index,” Arab. J. Sci. Eng., vol. 45, no. 8, pp. 7065–7076, Aug. 2020, doi: 10.1007/s13369-020-04718-w.

[3] [3] A. Atalan, “Application of Integer Response Optimization Models for the Healthcare Resources,” Acta Infologica, vol. 7, no. 1, pp. 81–93, Apr. 2023, doi: 10.26650/acin.1096774.

[4] [4] A. Atalan, H. Şahin, and Y. A. Atalan, “Integration of Machine Learning Algorithms and Discrete-Event Simulation for the Cost of Healthcare Resources,” Healthcare, vol. 10, no. 10, p. 1920, Sep. 2022, doi: 10.3390/healthcare10101920.

[5] [5] M. Maleki, N. Majlesinasab, and M. M. Sepehri, “Two new models for redeployment of ambulances,” Comput. Ind. Eng., vol. 78, pp. 271–284, Dec. 2014, doi: 10.1016/J.CIE.2014.05.019.

[6] [6] A. Atalan, “THE IMPACTS OF HEALTHCARE RESOURCES ON SERVICES OF EMERGENCY DEPARTMENT: DISCRETE EVENT SIMULATION WITH BOX-BEHNKEN DESIGN,” PONTE Int. Sci. Res. J., vol. 75, no. 6, pp. 12–23, 2019, doi: 10.21506/j.ponte.2019.6.10.

[7] [7] M. Farahani et al., “Impact of health system inputs on health outcome: A multilevel longitudinal analysis of Botswana national antiretroviral program (2002-2013),” PLoS One, 2016, doi: 10.1371/journal.pone.0160206.

[8] [8] R. Salway, R. Valenzuela, J. Shoenberger, W. Mallon, and A. Viccellio, “Emergency Department (Ed) Overcrowding: Evidence-Based Answers to Frequently Asked Questions.,” Rev. Médica Clínica Las Condes, vol. 28, no. 2, pp. 213–219, Mar. 2017, doi: 10.1016/J.RMCLC.2017.04.008.

[9] [9] N. R. Hoot et al., “Forecasting Emergency Department Crowding: A Discrete Event Simulation,” Ann. Emerg. Med., 2008, doi: 10.1016/j.annemergmed.2007.12.011.

[10] [10] T. Tenbensel, L. Chalmers, P. Jones, S. Appleton-Dyer, L. Walton, and S. Ameratunga, “New Zealand’s emergency department target - Did it reduce ED length of stay, and if so, how and when?,” BMC Health Serv. Res., vol. 17, no. 1, 2017, doi: 10.1186/s12913-017-2617-1.

[11] [11] M. Rahiminia, S. Shahrabifarahani, Z. Mojaradi, A. Aghsami, and F. Jolai, “A queueing-inventory model to control the congestion of patients and medical waste in the medical centers, a case study,” J. Manag. Anal., pp. 1–30, 2023.

[12] [12] A. Atalan and C. Donmez, “Employment of Emergency Advanced Nurses of Turkey: A Discrete-Event Simulation Application,” Processes, vol. 7, no. 1, p. 48, Jan. 2019, doi: 10.3390/pr7010048.

[13] [13] M. Yousefi, M. Yousefi, F. S. Fogliatto, R. P. M. Ferreira, and J. H. Kim, “Simulating the behavior of patients who leave a public hospital emergency department without being seen by a physician: a cellular automaton and agent-based framework,” Brazilian J. Med. Biol. Res., vol. 51, no. 3, Jan. 2018, doi: 10.1590/1414-431x20176961.

[14] [14] A. Atalan and C. C. Dönmez, “Optimizing experimental simulation design for the emergency departments,” Brazilian J. Oper. Prod. Manag., vol. 17, no. 4, pp. 1–13, 2020, doi: 10.14488/BJOPM.2020.026.

[15] [15] B. E. J. M. Driesen, B. H. G. van Riet, L. Verkerk, H. J. Bonjer, H. Merten, and P. W. B. Nanayakkara, “Long length of stay at the emergency department is mostly caused by organisational factors outside the influence of the emergency department: A root cause analysis,” PLoS One, vol. 13, no. 9, pp. e0202751–e0202751, Sep. 2018, doi: 10.1371/journal.pone.0202751.

[16] [16] S. Srinivas, R. P. Nazareth, and M. Shoriat Ullah, “Modeling and analysis of business process reengineering strategies for improving emergency department efficiency,” Simulation, vol. 97, no. 1, pp. 3–18, Jan. 2021, doi: 10.1177/0037549720957722.

[17] [17] S. Enayati, M. E. Mayorga, H. K. Rajagopalan, and C. Saydam, “Real-time ambulance redeployment approach to improve service coverage with fair and restricted workload for EMS providers,” Omega, vol. 79, pp. 67–80, Sep. 2018, doi: 10.1016/J.OMEGA.2017.08.001.

[18] [18] C. Duguay and F. Chetouane, “Modeling and Improving Emergency Department Systems using Discrete Event Simulation,” Simulation, 2007, doi: 10.1177/0037549707083111.

[19] [19] A. Goienetxea Uriarte, E. Ruiz Zúñiga, M. Urenda Moris, and A. H. C. Ng, “How can decision makers be supported in the improvement of an emergency department? A simulation, optimization and data mining approach,” Oper. Res. Heal. Care, vol. 1pp. 102–122, Dec. 2017, doi: 10.1016/J.ORHC.2017.10.003.

[20] [20] A. Atalan, “Effect of Healthcare Expendıture on the Correlatıon Between the Number of Nurses and Doctors Employed,” Int. J. Heal. Manag. Tour., vol. 6, no. 2, pp. 515–525, Jul. 2021, doi: 10.31201/ijhmt.949500.