

Determining Factors Influencing Length of Stay and Predicting Length of Stay Using Data Mining in the General Surgery Department

Samaneh Aghajani¹, Mehrdad Kargari^{1,*}

¹ Department of Industrial Engineering, Tarbiat Modares University, Tehran, IR Iran

*Corresponding Author: Mehrdad Kargari, Assistant Professor, Department of Industrial Engineering, Tarbiat Modares University, Tehran, Iran. Tel: +98-9132529084, Email: kargari@modares.ac.ir

Received: 3 Dec. 2015; Accepted: 16 Mar. 2016; Online Published: 28 May. 2016

Abstract

Background: Length of stay is one of the most important indicators in assessing hospital performance. A shorter stay can reduce the costs per discharge and shift care from inpatient to less expensive post-acute settings. It can lead to a greater readmission rate, better resource management, and more efficient services.

Objective: This study aimed to identify the factors influencing length of hospital stay and predict length of stay in the general surgery department.

Methods: In this study, patient information was collected from 327 records in the surgery department of Shariati Hospital using data mining techniques to determine factors influencing length of stay and to predict length of stay using three algorithms, namely decision tree, Naïve Bayes, and k-nearest neighbor algorithms. The data was split into a training data set and a test data set, and a model was built for the training data. A confusion matrix was obtained to calculate accuracy.

Results: Four factors presented: surgery type (hemorrhoid), average number of visits per day, number of trials, and number of days of hospitalization before surgery; the most important of these factors was length of stay. The overall accuracy of the decision tree was 88.9% for the training data set.

Conclusions: This study determined that all three algorithms can predict length of stay, but the decision tree performs the best.

Keywords: Data Mining, Decision Tree, General Surgery, Length of Stay

1. Background

Length of Stay (LOS) is defined as the number of days that a patient is hospitalized in a hospital [1]. It is an actual parameter used in identifying the use of health resources [2]. Predicting LOS for an inpatient is a challenging but essential task for the operational success of a hospital. Since hospitals are faced with severely limited resources, including beds to hold admitted patients, predicting LOS will be useful to hospital administrators for more effective hospital resource planning and management. There has been considerable interest in controlling hospital costs. Therefore, LOS is considered an important measure of healthcare utilization and a determinant of hospitalization costs. Clinical trials, electronic patient records, and computer-supported disease management will increasingly produce huge amounts of clinical data. LOS reflects physicians' decisions for patients to remain in hospital or not. The social problems, lack of services, lack of facilities, fault detection devices and other problems of a patient can increase his/her length of stay [3]. It is difficult for a manager to predict LOS, but it is essential and useful. In addition, many hospitals cannot predict or measure future admission requests [4]. If the inpatient LOS is predicted efficiently, the planning and management of hospital resources can be greatly enhanced. The purpose is to identify patterns affecting LOS that may help reduce costs. In recent years, the term 'data mining' has been increasingly used in medical literature; however, has been little research on predicting length of stay [5].

Since the evaluation of hospital performance using key indicators is an active hospital assessment methods, many studies calculate and compare LOS indicators to achieve

better performance. Arab et al. determined the factors affecting LOS in public hospitals in Lorestan Province, Iran. In their work, they demonstrated that age, gender, marital status, place of residence, occupation, type of referral, type of insurance, reason for admission, and discharge status impact the average length of stay. The researchers used T-test and one-way ANOVA in their study [6]. In another study, risk factors for prolonged stay after colorectal surgery were identified. That survey showed that congestive heart failure, high levels of albumin, and anemia are three factors affecting LOS in patients after colorectal surgery [7]. Hosseini et al. showed that number of beds, sex, age, and type of insurance did not increase or decrease LOS in hospital [8]. Huang et al. determined the factors associated with LOS for pneumonia patients. The authors' population-based prospective cohort study investigated 2,757 adults who were admitted over a two-year period. In this work it was shown that age, functional status, co-morbid disease, and weight were factors affecting LOS [9]. Some studies have shown a positive relationship between length of stay and hospital capacity and rate of unwelcome reception. In their study, Holland et al. showed a direct relationship between diagnostic laboratory quality and LOS in the emergency department [10]. Xiao et al. studied data from a medical hospital and showed that shock and the need for nutritional support have a significant relationship with LOS in the intensive care unit [11]. Robinson et al. conducted a survey of experts to propose a predictive model. Their study showed that the total number of new entries per day can be predicted 3 or 4 days in advance [5]. Bahrami et al. predicted the mortality rate of patients and LOS in the intensive care unit using the APACHE IV

method [12]. Rowan et al. used an artificial neural network to predict LOS for a cardiac surgical intensive care unit patient [13]. Li et al. used a backward propagation neural network model to predict LOS with 80% accuracy. In their study, five predictors were identified: days hospitalized before operation, wound grade, operation approach, charge type, and number of admissions [14]. Liu et al. extracted 15 features, including demographic characteristics, reasons for admission, registration details, output and length of stay from patient data. Then they applied two successful and widely-used classifiers after pre-processing [15]. In another study, the researcher used data mining techniques to investigate LOS for children incubated in the United States and Egypt. Logistic regression, Bayesian classifier, and support vector machine were applied to the data. The results showed the support vector machine was the best fit [16]. Tanuja et al. used elderly hospital electronic discharge data to predict LOS. They used four methods, namely the artificial neural network method, the Bayesian classifier, the K-nearest neighbor algorithm (K-NN), and the W-J48 decision tree. The results showed that MLP performed better than the other methods [17]. Rezai et al. studied 4,948 cases to determine and predict the patient's LOS. In this study, length of stay was divided into three classes, and the C5.0 decision tree algorithm, the SVM algorithm, and the neural network were used to predict it. SVM was chosen as the best predictor [4]. Gomez et al. used clustering techniques, decision trees, association rules, and OLAP (On Line Analytical Processing) to predict LOS. Their results showed that LOS is longer for men than for women. They also found that disease severity contributes to LOS [18].

1.1. Data Mining Algorithms

Data mining can be used to discover information and hidden patterns within a database. This new information can be used to improve actions and procedures. Today, data analysis is critical for medical decision making and management. Data mining is used in the medical field for applications such as disease prediction, disease diagnosis, remedy type determination, healthcare electronic records, hospital infection control, and hospital gradation [19, 20]. As mentioned above, data mining algorithms have been successfully applied to predict LOS. The current study used some of the most common predictive data mining methods as described in the following.

1.1.1. Decision Tree

The decision tree model is useful and popular for classification [21]. Classification algorithms require that the classes be defined based on attribute values. A decision tree is a graphic representation in the form of a tree from obtained knowledge and is presented in the form of nodes, branches, and leaves [22]. In a decision tree, nodes and branches are arranged hierarchically, which makes it easy to understand and interpret [4]. A decision tree is constructed in several recursive steps: 1) choose an attribute to put at the root node; 2) Make one branch for each possible value; 3) Repeat the process to construct the tree [21]. A C5.0 decision tree is the most common decision tree algorithm. In this study, researchers used a W-j48 decision tree, which is a Java implementation of the C4.5 decision

tree.

1.1.2. Naive Bayes Classification

The Naive Bayesian classifier is a simple and efficient classifying algorithm that performs better compared with other data mining techniques such as decision tree and neural network algorithms [23]. The Naive Bayesian classifier is suited when the dimensionality of the inputs is high. This derives the posterior probability $P(C|X)$ from the prior probability $P(C)$ and $P(X)$ and the conditional probability $P(X|C)$ by using the following relation [14]:

$$(1) \quad P(C|X) = (P(X|C) \cdot P(C)) / P(X)$$

1.1.3. K-Nearest Neighbor (KNN)

K-NN is a supervised learning algorithm that determines the k-neighborhood parameter at the first [22]. It is the simplest machine-learning algorithm. K-NN is a type of instance-based learning. It uses distance measurements such as Euclidean, Hamming, and Manhattan to calculate the distances of samples from each other. This lazy learning searches slowly for the most similar instances and detects a complete training set at the time of classification [23].

2. Objective

Data mining techniques can be implemented retrospectively on massive amounts of data in an automated matter. Data mining algorithms have also been successfully applied to predict LOS. One issue considered by researchers was to provide an efficient and accurate model to predict LOS. Therefore, this study attempted to predict date of discharge and LOS because these values can facilitate scheduling elective admissions and reduce variance bed occupancy [5]. Then, data mining techniques were applied to extract useful knowledge and suggest a model to estimate LOS for general surgery patients in Shariati Hospital.

3. Methods

3.1. Patient Population

Between March 21, 2013 and June 21, 2013, 327 patients underwent general surgery in the Department of General Surgery in Shariati Hospital, Tehran, Iran. As this dataset is essentially an audit and is reported in a global manner without identifying individual patients, this study complies with ethical issues.

3.2. Data Collection

The data sets used in this study were stored in the hospital information system (HIS). A new data set was extracted and constructed for LOS in General Surgery. Some attributes were collected from the HIS, and the rest were extracted from patient records. Features with acceptable classes and values are provided in Appendix 1. The data set contained 30 attributes. Finally, the data set was split into two groups, categorical features and numerical features. Researchers attempted to use the patient information available at time of admission and the discharge summary details to develop a model to predict LOS. Unavailable data and files were removed. Ultimately, 327 records were selected for further analysis.

3.3. Data Preprocessing

For data mining, clinical data should be extracted from dedicated databases that have been purposely collected to study a particular clinical problem. The hospital data set

had many features with missing values. Preprocessing is essential to achieving optimal results. The following cleansing and preprocessing activities were performed: repeated records, fields with spelling errors, additional tokens, other irregularities, and irrelevancies were deleted [4]. Missing values in patient data are common in medical environments. Since the required data was accurately collected by the doctors and nurses, the problem of outlier or missing data was negligible.

3.4. Feature Selection

The first step in developing any classification solution is to identify the independent input variables that contribute to the classification decision. This study reviewed the literature to extract useful and appropriate features. Then, features extracted from previous research and the researcher’s proposal were given to experts to be endorsed. Four specialists in general surgery were surveyed. Features were selected if the experts believed it to be effective on

LOS. The number of votes to each feature was counted and obtained approval for each feature. Percent confirmed of the experts for any characteristics is given in Table 1.

3.5. Attribute Coding

Data was coded by valid resources, such as vital signs associations and expert opinion. Scaling and coding features are given in Appendix 1.

3.6. Training and Test Data Sets

After cleaning and preprocessing, 327 completed records were extracted and obtained for data mining tasks. Separating the data into training and testing sets is an important part of evaluating data mining models. In this study, 70% of the data was used for training, and 30% was used for testing. The training set was used to adjust the parameters of the models, and the testing sets were used to evaluate its predictive ability.

Table 1. Experts’ questionnaire

Attribute	Expert 1	Expert 2	Expert 3	Expert 4	Result
1 Age	√	√	√	√	100%
2 Gender	√			√	50%
3 Marital status	√			√	50%
4 Lodging	√	√	√	√	100%
5 Insurance type	√	√		√	75%
6 Smoking history	√		√		50%
7 Pulse rate	√			√	50%
8 Systolic blood pressure	√	√	√	√	100%
9 Diastolic blood pressure	√	√	√	√	100%
10 Co-morbidities (hypertension, diabetes)	√		√	√	75%
11 Hospitalization / surgical history	√			√	50%
12 Hemoglobin	√	√	√	√	100%
13 Fever before surgery	√			√	50%
14 Fever after surgery	√	√	√	√	100%
15 Respiratory rate before surgery	√			√	50%
16 Respiratory rate after surgery	√		√	√	75%
17 Patient's condition at time of admission	√			√	50%
18 Patient's condition at time of discharge	√			√	50%
19 Surgery type	√		√	√	75%
20 Length of surgery	√		√	√	75%
21 Number of surgeries	√		√	√	75%
22 Using ICU	√	√	√	√	100%
23 Average number of visits per day				√	25%
24 Transmission between parts	√	√	√	√	100%
25 Preoperative hospitalization	√				25%
26 Number of tests	√		√	√	75%
27 Interval between first and last medical consultation	√		√	√	75%
28 Number of consultations	√	√		√	75%
29 Admission type	√	√		√	75%
30 Interval between discharge order and discharge	√	√		√	75%

4. Results

4.1. Statistical Analysis

The average age of patients presenting for surgery was 45.04 years, and the study group consisted of 65.8% male and 34.2% female patients. The average number of examinations was 1.5 per day, and the average length of stay before surgery was 2.72 days. In this study, LOS was divided into three classes as given in Table 2.

Table 2. Categorizing LOS

Class	LOS range (day)	Frequency	Relative frequency
1	1 <= los <= 3	122	37.3%
2	4 <= los <=5	66	20.1%
3	6 <= los	139	42.2%

These data mining models were developed by a data mining classification tool. The model was created using

training data and evaluated using test data. Then the models were compared against the accuracy criterion. Rapid miner was used to build mining models, and three algorithms were used to build the model: KNN, Bayesian classification, and w-j48 decision tree. Each confusion matrix is given in Table 3.

For applicability of the method in the hospital setting, data upon admission was used to build prediction models. In this stage, all information collected after admission was removed. Then features such as demographic characteristics, co-morbidities, history, and diagnosis were used to apply the prediction model. The decision tree algorithm was used to build the prediction model. The accuracy of the obtained model was poor, but it is worthwhile for use in hospitals. The fusion matrix for the decision tree is given in Table 4.

4.2. Post Operation Prediction

In this work only pre-operation information was used to predict post-operation length of stay. All features associated with the postoperative situation were excluded and those remaining were used to

make predictions using the decision tree. Finally, the decision tree was obtained using 0.7 training data and 0.3 test data. This decision tree can predict post-operation LOS with 84.69% accuracy. The confusion matrix for the decision tree is given in Table 4.

Table 3. Confusion matrix for KNN- naïve Bayes- Decision tree W-J48

	Class 1 true			Class 2 true			Class 3 true			Class precision		
	KNN	Bayesian classification	Decision tree	KNN	Bayesian classification	Decision tree	KNN	Bayesian classification	Decision tree	KNN	Bayesian classification	Decision tree
Class 1. Pred	30	36	32	2	4	3	3	4	2	68.75%	81.82%	86.84%
Class 2. Pred	2	0	0	11	9	7	3	1	1	89.74%	90%	87.50%
Class 3. Pred	0	2	2	4	1	2	35	40	40	85.71%	93.02%	90.91%
Class recall	93.75%	94.74%	94.29%	64.71%	64.29%	58.33%	85.37%	88.89%	93.02%			

Table 4. Fusion matrix for Decision tree W-J48 in admission time for the post-operation length of stay

	Class 1 true		Class 2 true		Class 3 true		Class precision	
	accuracy: 75.25%	accuracy: 84.69%						
Class 1. Pred	44	73	9	6	10	1	69.84%	91.25%
Class 2. Pred	0	1	0	0	0	1	0	0
Class 3. Pred	1	3	5	3	32	10	84.21%	62.5%
Class recall	97.78%	94%	0	0	76.19%	83.33%		

4.3. Clustering

Data was clustered using the k-means algorithm which was run 500 times; different values were given for the number of clusters (k = 2, 3, 4, 5). The optimal number of clusters was three using the Davies-Bouldin Index. The clustering results showed the distribution of LOS in the clusters as given in Table 5.

4.4. Important Features

Each data mining model uses a subset of features in the classification model. In this study, the effects of the features on LOS were determined using linear regression and with the SPSS tool. The features are shown in Table 6.

4.5. Data Visualization

In this section, a five-dimension chart is built using 4

important constitutive features in the decision tree. This chart is given in Figure 1, APPENDIX 2. As seen in this chart, the four features of hemorrhoids, number of tests, pre-operation length of stay, and number of visits per day were selected as independent variables and LOS was selected as the dependent variable. There is a direct relation between the number of tests, pre-operation length of stay, and LOS and a negative relation between number of visits per day, hemorrhoid surgery, and LOS.

Table 7 shows that post-operation hospitalization was indicated as an important feature with two prediction models, but it obtained the least number of expert votes. Moreover, number of visits per day was indicated as a more effective factor with three algorithms, but experts did not recognize it as an important factor.

Table 5. Clustering results

Cluster 3	Percent of frequency			Frequency			Length of stay (days)
	Cluster 2	Cluster 1	Cluster 3	Cluster 2	Cluster 1		
0	100%	0	0	16	0	1	
2%	98%	0	1	52	0	2	
2%	98%	0	1	55	0	3	
56%	94.4%	0	2	34	0	4	
29.5%	70.5%	0	10	24	0	5	
31.6%	68.4%	0	6	13	0	6	
43.5%	56.5%	0	10	13	0	7	
58.3%	41.7%	0	7	5	0	8	
41.2%	58.8%	0	7	10	0	9	
70%	30%	0	7	3	0	10	
50%	25%	25%	2	1	1	11	
33%	33%	33%	2	2	2	12	
58%	14%	28%	4	1	2	13	
100%	0	0	1	0	0	14	
20%	20%	60%	1	1	3	15	
67%	0	33%	4	0	2	16	
0	33%	67%	0	1	2	17	
24%	13%	62%	7	4	18	18 to 68	

Table 6. Important features

Feature	KNN	Bayesian classification	Decision tree
Abdominal surgery	*		
Endocrinology surgery	*	*	
Hemorrhoid surgery	*		*
Average number of visits per day	*	*	*
Number of medical consultations	*		
Gastroenterology surgery		*	
Number of surgeries		*	
Interval between discharge order and discharge		*	
Transmission between parts		*	
Preoperative hospitalization		*	*
Number of tests			*

Table 7. Feature comparison between data mining algorithms and experts' opinions

Feature	KNN	Bayesian classification	Decision tree	Expert support (%)
Surgery type	*	*	*	75%
Average number of visits per day	*	*	*	25%
Number of medical consultations	*			75%
Number of surgeries		*		75%
Interval between discharge order and discharge		*		75%
Transmission between parts		*		100%
Preoperative hospitalization		*	*	25%
Number of tests			*	75%

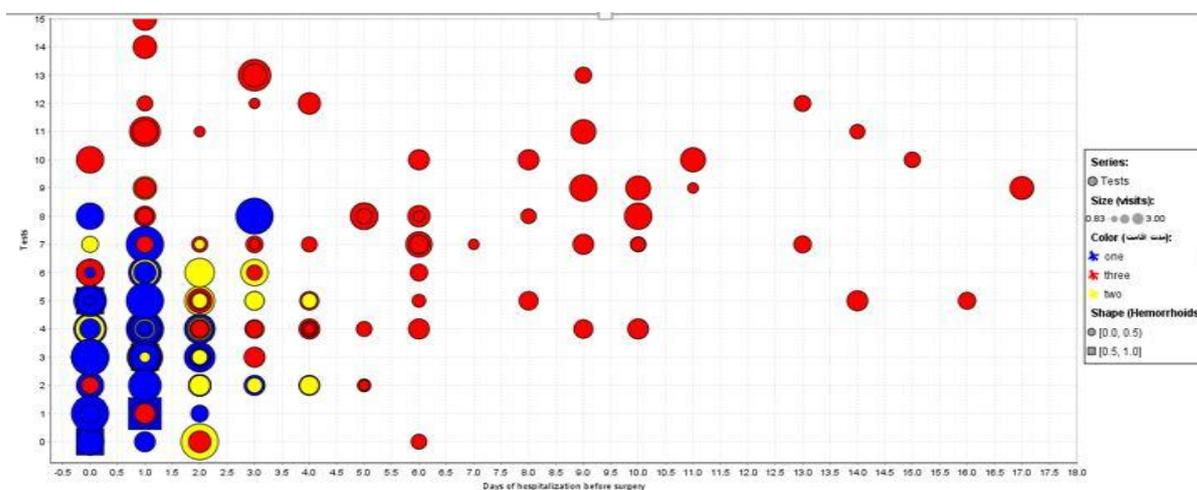


Figure 1. Five-dimension chart

5. Discussion

This study investigated the determinants of length of hospital stay in patients admitted to a general surgery department. The findings show that surgery type, number of surgeries, interval between discharge order and discharge, transmission between parts, average number of visits per day, number of medical consultations, preoperative hospitalization, and number of tests are important factors affecting length of stay. Among the three data mining algorithms compared in this study, the decision tree model was found to be the best predictor. Afterwards, all the features were used to obtain these prediction models. Many studies of length of hospital stay predict the duration of stay based on laboratory parameters or other quantifiable variables [5]. The current findings showed that a LOS greater than 6 days is associated with preoperative hospitalization and number of tests. There was a significant tendency for LOS to be shorter in patients with hemorrhoids and an average number of visits per day of more than two. This study also demonstrated that co-morbidities do not significantly influence LOS, while the average number of visits has an inverse effect on LOS. Some studies have reported that patient demographics and attributes were the two major factors contributing to identifying patient LOS [6]. In this study,

the extracted rule demonstrated that patients with hemorrhoids, less than five tests, and a preoperative hospitalization of less than two days and had a normal LOS in hospital. Marital status was not significantly related to LOS. While the promising results of this study indicate that Naïve Bayes and KNN models can predict a patient's LOS, the decision tree model has the best fit and is optimal for predicting the LOS of general surgery patients. In addition to disease-related factors, LOS may be affected by factors unrelated to the disease, such as availability of hospital physicians and rehabilitation facilities as well as discharge possibilities. In this study, expert opinions were used and were compared with the findings of this study to understand differences.

Table 7 shows that pre-operation hospitalization was indicated as an important feature with two prediction models, but it obtained the least number of expert votes. Number of visits per day was also indicated as a more effective factor with three algorithms, but the experts didn't recognize it as an important factor. Although they selected age, lodging, blood pressure, hemoglobin, post-operation fever, and using ICU as important factors, none of these factors were determined to be significant using data mining algorithms. These differences show that medical experts don't have enough insight to determine

what factors influence LOS. Thus, new methods should be used to understand invisible patterns in data. These finding can help hospitals improve activities recognized with data mining algorithms.

6. Conclusion

Length of stay is one of the most important indicators in assessing hospital performance. It can facilitate scheduling elective admissions and reduce variance in bed occupancy [5]. Using data mining algorithms can help predict LOS in hospital. Some preoperative features are used for making predictions. These features were selected through a feature selection method and then used with three data mining algorithms, of which the decision tree was found to be the best.

Acknowledgments

This study was a part of a master's thesis supported by the Industrial Engineering Department of Tarbiat Modares University. The authors would like to thank Dr. Soroush (CEO of Hospital), Dr. Siamak Kargar (Medical Consultant), and other staff members of the Information Tech Unit of Shariati Hospital & Research Center for allowing us to use their patient data and for their assistance in retrieving it.

Authors' Contributions

All authors contributed equally in the preparation of this paper.

Conflict of Interest

No conflict of interest has been declared by the authors.

References

1. Yaghoubi M, Karimi S, Ketabi S, Javadi M. Factors affecting patients' length of stay in Alzahra hospital based on hierarchical analysis technique. *Health Inf Manage.* 2011;8(3):326-34.
2. Lim A, Tongkumchum P. Methods for analyzing hospital length of stay with application to inpatients dying in Southern Thailand. *Glob J Health Sci.* 2009;1(1):27. [DOI]
3. Jack W. Principles of health economics for developing countries: World Bank Publications; 1999.
4. Hachesu PR, Ahmadi M, Alizadeh S, Sadoughi F. Use of data mining techniques to determine and predict length of stay of cardiac patients. *Health Inform Res.* 2013;19(2):121-9. [DOI]
5. Robinson GH, Davis LE, Leifer RP. Prediction of hospital length of stay. *Health Serv Res.* 1966;1(3):287.

6. Arab M, Zarei A, Rahimi A, Rezaiean F, Akbari F. Analysis of factors affecting length of stay in public hospitals in Lorestan Province, Iran. *Hakim Res J.* 2010;12(4):27-32.
7. Lobato LFdC, Ferreira PCA, Wick EC, Kiran RP, Remzi FH, Kalady MF, et al. Risk factors for prolonged length of stay after colorectal surgery. *J Coloproctol.* 2013;33(1):22-7. [DOI]
8. Hosseini AF, Gohari MR. Application of multilevel model in determining the effective factors in the length of stay among appendectomy patients. *Razi J Medl Sci.* 2014;20(115):70-7.
9. Huang JQ, Hooper PM, Marrie TJ. Factors associated with length of stay in hospital for suspected community-acquired pneumonia. *Can Respir J.* 2006;13(6):317. [DOI]
10. Holland LL, Smith LL, Blick KE. Reducing laboratory turnaround time outliers can reduce emergency department patient length of stay. *Am J Clin Pathol.* 2005;124(5):672-4. [DOI]
11. Xiao J, Douglas D, Lee AH, Vemuri SR. A Delphi evaluation of the factors influencing length of stay in Australian hospitals. *Int J Health Plann Manage.* 1997;12(3):207-18. [DOI]
12. Bahrami N, Soleimani MA, Shraif Nia SH, Shaigan H, Masoodi R, Ranjbar H. Predicted duration of hospital stay and percentage of mortality in patients intensive care unit with APACHE IV. *Urmia Med J.* 2012;23(5):466-70.
13. Rowan M, Ryan T, Hegarty F, O'Hare N. The use of artificial neural networks to stratify the length of stay of cardiac patients based on preoperative and initial postoperative factors. *Artif Intell Med.* 2007;40(3):211-21. [DOI]
14. Li J-S, Tian Y, Liu Y-F, Shu T, Liang M-H. Applying a BP neural network model to predict the length of hospital stay. *Health Information Science: Springer;* 2013. p. 18-29.
15. Liu P, Lei L, Yin J, Zhang W, Naijun W, El-Darzi E. Healthcare data mining: predicting inpatient length of stay. 2006.
16. Hagar Fady S, Taha E-ST, Mervat Mahmoud M. Applying Data Mining Technique for Predicting Incubator Length of Stay in Egypt and USA. *Int J Comput Appl.* 2013;71(8):21-8.
17. Tanuja S, Acharya DU, Shailesh K. Comparison of different data mining techniques to predict hospital length of stay. *J Pharm Biomed Sci.* 2011;7(7).
18. Gómez V, Abásolo JE. Using data mining to describe long hospital stays. *Paradigma.* 2009;3(1):1-10.
19. Khalilinezhad M, Minaee Bidgoli B, editors. Clinical data mining. Proceeding of the 3rd Iran Data Mining Conference; 2010.
20. Obenshain MK. Application of data mining techniques to healthcare data. *Infect Control Hosp Epidemiol.* 2004;25(08):690-5. [DOI]
21. Liu P, El-Darzi E, Vasilakis C, Chountas P, Huang W, Lei L. Comparative analysis of data mining algorithms for predicting inpatient length of stay. *PACIS 2004 Proceedings.* 2004:86.
22. Kaya E, Findik O, Babaoglu I, Arslan A. Effect of discretization method on the diagnosis of Parkinson's disease. *Int J Innov Comp Inf Control.* 2011;7:4669-78.
23. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. *Int J Med Inform.* 2008;77(2):81-97. [DOI]

Appendix 1

Attribute	Coding
Age	Numerical
Gender	Female (0) Male (1)
Co-morbidities	Yes (1) No (0)
Number of surgeries	Numerical
Preoperative hospitalization	Numerical
Marital status	Single (0) Married (1)
Lodging	Native (1) Other (0)
Insurance type	Without insurance (0) Social secure insurance (1) villager services insurance (2) personal insurance (3) government staff insurance (4) army insurance (5) other (6)
Smoking history	Yes (1) No (0)
Patient's condition at time of admission	Low risk (0) Medium risk (1) High risk (2)
Patient's condition at time of discharge	Recuperation (0) Own acquiescence (1) Death (2)
Admission type	Emergency (0) Elective (1)
Using ICU	Yes (1) No (0)
Surgery type	Yes (1) No (0)
Hospitalization/surgical History	Yes (1) No (0)
Preoperative/postoperative fever	Yes (1) No (0)
Preoperative/postoperative respiratory rate	Ok (1) not Ok (0)
Length of surgery	Numerical
Number of tests	Numerical
Number of consultations	Numerical
Interval between first and last medical consultations	Numerical
Interval between discharge order and discharge	Numerical
Length of stay	Short (class 1) Medium (class 2) Long (class 3)