INTERNATIONAL JOURNAL OF SOFTWARE ENGINEERING & COMPUTER SYSTEMS (IJSECS)
ISSN: 2289-8522 e-ISSN: 2180-0650
VOL. 7, ISSUE 2, 64 – 73
DOI: https://doi.org/10.15282/ijsecs.7.2.2021.7.0090

ORIGINAL ARTICLE

Redesigning Post-Operative Processes Using Data Mining Classification Techniques

H. Ghazi Alwattar[1]

[1]Department of Computing, University of Worceter, Henwick Grove, WR2 6AJ, UK.

ABSTRACT – Data mining classification models are developed and investigated in this paper. These models are adopted to develop and redesign several business processes based on post-operative data. Post-operative data were collected and used via the Waikato Environment for Knowledge Analysis (WEKA), to investigate the factors influencing patients' admission after surgery and compare the developed DM classification models. The results reveal that each implemented DM technique entails different attributes affecting patients' post-surgery admission status. The comparison suggests that neural networks outperform other classification techniques. Further, the optimal number of beds required to accommodate post-operative patients is investigated. The simulation was conducted using queuing theory software to compute the expected number of beds required to achieve zero waiting time. The results indicate that the number of beds required to accommodate post-surgery patients waiting in the queue is the length of 1, which means that one bed will be available due to patient discharge.

ARTICLE HISTORY
Received: 20 Mac 2021
Revised: 23 Sept 2021
Accepted: 5 Oct 2021

KEYWORDS
Data mining
Data modelling and simulation
Decision making
Neural networks
Bayes' networks
Healthcare management

INTRODUCTION

Data mining has become a research area with increasing importance. Organizations of all sizes have started to develop and deploy data mining technologies to leverage data resources to enhance their decision making capabilities [1]**.** The capabilities of data analysis are a crucial factor of success for business today. The Main parts of the business world are based on IT and deal with huge amounts of electronic data. Approaches to modeling data mining in business contexts are limited in addressing a major current trend. Data Mining becomes more and more an integral part of executing a business. Tasks like placing advertisements, recommending products, or detecting fraud have become standard application yields of data mining, and have a serious implication for business products. A frequent re-engineering of business processes is a consequence of this development [2].

Data mining (also known as database exploration and knowledge discovery) is the process of extracting hidden patterns from large databases and using them to build predictive models [3]. This method of extracting concealed knowledge from large datasets is effective in solving business and scientific problems and obtaining a competitive advantage. Traditional data analysis methods, largely based on humans working directly with data, simply do not scale up to handle large data. Big data in healthcare institutions has led to an urgent need for new techniques and resources that can automatically turn processed data into valuable information and knowledge [4], [5]. In the healthcare sector, several ways of knowing how to gather, store and prepare data for mining are available. A few such methods include standardization of clinical terminology and the exchange of data across organizations to improve the value and utilization of healthcare data through mining applications [6].

The business process of a high bed occupancy rate, poses several challenges for hospital management. Variation in the arrival of surgical patients and their duration of stay can also lead to a discrepancy between the demand and supply of staffed beds. The effects of this imbalance are compounded during times of high bed occupancies. Patient care is potentially compromised when demand exceeds the supply of staffed beds [7].

The goal of the present study is threefold. First, we propose the use of DM classification techniques to better support the redesigning process of some hospital management business processes by retrieving the required information from large datasets recorded by a hospital via DM models. The redesigning process is based on post-operative data which affects the management side of things in the sector, such as bed occupancy rate and human resources in hospitals. In order to build these DM models, different algorithms are proposed such as neural networks and Bayes' decision tree and rule Induction methods. These models enable hospitals to reduce their costs by automating hospital management functions in determining the number of beds and human resources in the hospital [8].

Second, every classification algorithm is implemented at a time using WEKA based on post-surgery patient status attributes, and the resultant models are compared and tested to determine which one has the most influence on whether a patient will stay after surgery or check out from the hospital, eventually affecting the bed occupancy and human resources [9].

*CORRESPONDING AUTHOR | H. Gahzi Alwattar | ✉ haider_gk@yahoo.com

© The Authors 2021. Published by Penerbit UMP. This is an open access article under the CC BY license.

Third, simulation experiments were conducted using queuing theory to evaluate the optimal number of beds required to accommodate post-operative patients. These experiments were based on modeling the number of patients expected to stay after surgery (arrival) and the number of days they are expected to stay (service time) as statistical distribution based on real data. The goal is to compute the expected number of beds required to achieve a strategy of zero waiting time.

The paper is organized as follows. Section 2 describes the data and research methodology; Section 3 presents the selected classification techniques implemented via WEKA software to build data mining models; Section 4 discusses the Queueing Theory software used to determine the beds needed for post-surgery patients, and the conclusion and suggestions for future work are presented in Section 5.

RELATED WORK

Works on DM techniques in healthcare are related to and motivated by studies in the data analysis area in general and DM techniques in particular. These studies appeared mostly in the last decade. In 2010, one of the first studies was released [10], which deals with healthcare administration and specifically covers the issues of operating room planning and waiting list management. The paper analyses whether the Master Surgery Schedule (MSS) may be changed to better accommodate shifting surgery demand using an optimization-based method. Second, an expanded optimization-based strategy is described, which includes postoperative beds, in which alternative policies linked to priority rules are simulated to show their impact on average waiting time [10]. The findings suggest that implementing a specific strategy reduces average patient waiting time and operation cancellations while increasing operating room utilization.

Agent-Based Simulation (ABS) appears to be a potential approach to many challenges involving simulating complex systems of interacting entities, according to the applications covered in the paper. Furthermore, current ABS tools and platforms are rarely used, and simulation software is instead created from the ground up using a standard programming language [11]. A recent study looked at how classification-based data mining approaches like Rule-based, Decision tree, Nave Bayes, and Artificial Neural network may be applied to large amounts of healthcare data [11]. They investigated the topic of limiting and summarising alternative data mining methods as part of their research. They also concentrated on predicting combinations of numerous target attributes using various techniques. They provided an innovative and effective heart attack prediction approach utilizing data mining in their study [12].

In 2009, a survey of current KDD strategies for healthcare and public health utilizing data mining technologies was released [13]. It goes through some of the most important topics and challenges surrounding data mining and healthcare in general. The study discovered a rising number of data mining applications, including health care centre analysis for better health policy-making, disease outbreak identification and preventable hospital fatalities, and the detection of fraudulent insurance claims. In general, the study aims to give a survey of current data mining approaches for knowledge discovery in databases that are used in medical research and public health today.

The results of this study can help us realize that, before commencing ~~on~~ data mining, a company must first establish clear policies on patient privacy and security. It is responsible for enforcing this policy with its partners, as well as its branches and agencies. Rapid pandemic outbreaks, the need to detect illness starts in a non-invasive, painless manner, and the need to be more responsive to consumers – all of these factors contribute to a growing requirement for health companies to integrate data and use data mining to analyze these data sets [13]. Another recent survey of different DM techniques were provided such as classification, clustering, association, and regression in the healthcare sector [14]. Several examples of each one of these DM techniques was introduced and investigated based on their accuracy and performance. For example, the performance of classification techniques is affected by noisy data. Therefore, the success of using data mining starts with the availability of clean data. Other important issues that were discussed are the importance of patients' data privacy and how to use security measures to protect patients' data from being accessed by authorized users. Finally, the survey also highlighted the importance of accuracy in the prediction of the disease which can be accomplished by combining different DM techniques [14].

RESEARCH METHODOLOGY AND DATASETS

Data mining classification algorithms were implemented via the WEKA software to determine the factors affecting the post-surgery stay duration of a patient. Patient data were obtained from the UCI Machine Learning Repository [15]. The number of beds that need to be made available to accommodate post-operative patients and the waiting time for each patient in the queue were determined using data from 2–3 months of hospital records of daily operations. The data were analyzed using statistical analysis and ~~a~~ queuing simulation software.

Data mining classification techniques and the factores effecting patients staying after a surgery

In this section, the use of decision tree J48, rule induction PART, Bayes' net, and neural networks algorithms multilayer perceptron is described. These were implemented via the WEKA software, which is used to build data mining models to predict post-surgery patient status. The WEKA software is a common suite of machine learning algorithms for data mining tasks [16]. WEKA's native method of storage is in the ARFF format. A conversion was therefore carried out to make the test data accessible for analysis via WEKA. It is a free software licensed under the GNU General Public License [17]. The developed data mining models were used to predict the possibility of a patient either staying after surgery or going home based on several attributes. The main goal is to determine the attribute of the dataset which has the most effect on this decision of a patient via data mining models. This decision affects the management body of a

hospital by creating a need to offer more services to these patients such as providing more beds, nurse care, and nurse shifts.

Factors effecting patients staying after a surgery

The dataset was collected from the UCI Machine Learning Repository for post-operative patients. It was pre-processed, cleaned, and prepared for usage to build classification models. These models are developed based on the following attributes of post-operative patients:

1. L-CORE (internal temperature in C): high (> 37), mid (>= 36 and <= 37), low (< 36).

2. L-SURF (surface temperature in C): high (> 36.5), mid (>= 36.5 and <= 35), low (< 35).

3. L-O2 (oxygen saturation in %): excellent (>= 98), good (>= 90 and < 98), fair (>= 80 and < 90), poor (< 80).

4. L-BP (last blood pressure measurement): high (> 130/90), mid (<= 130/90 and >= 90/70), low (< 90/70).

5. SURF-STBL (surface temperature stability): stable, mod-stable, unstable.

6. CORE-STBL (core temperature stability) stable, mod-stable, unstable.

7. BP-STBL (blood pressure stability) stable, mod-stable, unstable.

8. COMFORT (patient's perceived comfort at discharge, measured as an integer between 0 and 20).

9. Decision ADM-DECS (discharge decision): I (patient sent to Intensive Care Unit), S (patient prepared to go home), A (patient sent to general hospital floor).

WEKA is used to build classification models based on the above-mentioned attributes using 50 records of post-operative patent datasets by implementing the selected classification techniques and their algorithms mentioned earlier. Following this, 40 records of post-operative patients' datasets were used to test the created models which required ensuring that these algorithms and models work properly and fulfill their potential. For the current paper, 50 records of post-operative patients were collected and used via WEKA to build classification models based on the above-mentioned attributes. The purpose of building these classification models is to identify the factors that determine whether a patient stays or leaves after surgery. After the implementation of each classification technique, the results of patients staying and leaving after surgery were computed in percentage.

Decision tree

Algorithm J48 was implemented via WEKA using 50 post-operative patients' dataset [18]. The results presented at the end of this section compare the implemented classification techniques using WEKA software (Table 2). We are concerned about the statistics representing the numbers of correctly and incorrectly classified instances. Except for the Kappa statistic, the remaining statistics determine specific error quantities based on the class probabilities assigned by the tree. In the present paper, the confusion matrix is considered for each class and how instances from that class receive various classifications. For class "b", 2 correctly classified instances were obtained but 14 was put into class "C", while for class "C", 33 were correctly classified. To determine which attributes most influence patients' decision to stay at a hospital after surgery, we consider the model of 50 records created by WEKA using the decision tree algorithm J48. The unpruned decision tree and the pruned decision tree illustrate how attributes are used by the classifier to make a decision. The leaf nodes show the class to which an instance is allocated if the node is achieved. The numbers in the brackets after the leaf nodes present the number of instances assigned to a node, followed by how many of those instances are incorrectly classified as a result. One attribute that has the most effect on the decision of a patient staying or leaving after surgery is COMFORT, which starts as the root of the tree.

Neural network

A neural network algorithm (multilayer perceptron) is used via WEKA software on post-operative patients' dataset of 50 records [19]. In the current paper, the concern is that the figures must represent the numbers of the correctly and incorrectly classified instances. Except for the Kappa statistic, the remaining statistics determine specific error quantities based on the class probabilities assigned by the neural network. Also, the confusion matrix is considered for each class and how instances from that class receive various classifications. For class "b", 15 correctly classified instances were obtained but 1 was put into class "c", while for class "C", 31 were correctly classified but 2 were put into class "b". For class "I", 1 was correctly classified.

To determine that exact attributes influencing the decision to stay or leave after surgery, the first created model of 50 records by WEKA was considered using the neural network algorithm, which is a multi-layer perceptron algorithm, as illustrated in Figure 1.

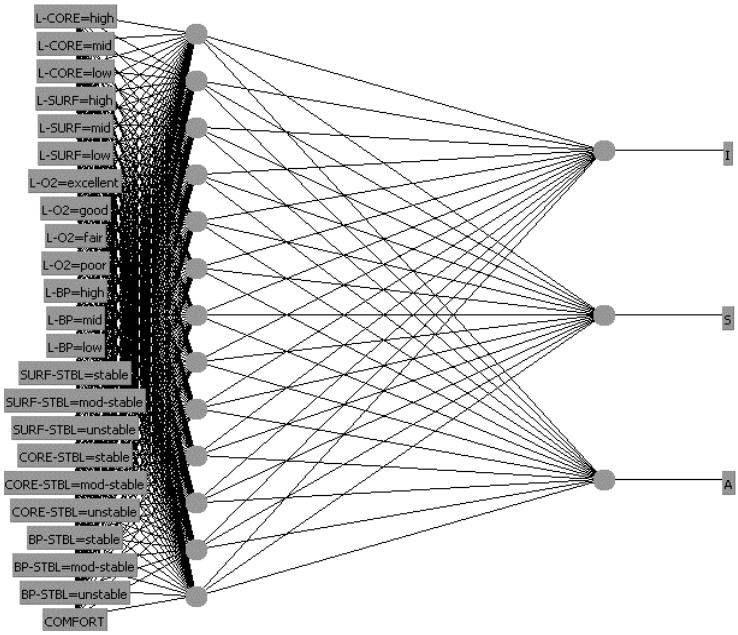

Figure 1. Neural network for post-operative patient's dataset.

Figure 1 illustrates the entire network with all its attributes to determine which attribute most influences the patient's decision for which attributes were isolated one at a time and the error rate was monitored. If the error rate is more than that of the neural network without isolating, it means this isolated attribute affects the decision. This procedure was carried out by implementing the neural network algorithm on a patient dataset of 50 records using WEKA software one at a time after isolating each attribute. The multi-layer perceptron is presented without isolating any attributes. The comparison of neural network algorithms with attribute isolation and without attribute isolation is listed in Table 1. Following this, the algorithm was implemented by isolating one attribute at a time to find out which ones have the most effect.

Table 1. Comparision between the used NN algorithm before and after attribute isolation

Attributes	L-BP	COMFORT	BP-STBL	L-CORE	L-SURF	L-O2	SURF-STBL	CORE-STBL
Error rate after Isolation	8%	8%	10%	10%	12%	6%	6%	6%
Error rate before Isolation	6%	6%	6%	6%	6%	6%	6%	6%

The above table shows the result of the used NN algorithm after isolating the attributes compared with the used NN algorithm before isolating any attributes, along with comparing the error rate of the implemented algorithm before the isolation with the implemented algorithm of each isolating attribute. It was found that the main attributes affecting the decision of a patient staying at a hospital are as follows:

1. L-SURF
2. L-CORE
3. BP-STBL
4. L-BP
5. COMFORT

The rest of the attributes were ignored as they don't have any effect on the patient's decision.

Rule induction

Rule induction algorithm (PART) was used via WEKA on post-operative patients' dataset of 50 records [20]. In the present paper, we are concerned with the number representing the correctly and incorrectly classified instances. Except for the Kappa statistic, the remaining statistics determine specific error quantities based on the class probabilities assigned by the rules. Also, the confusion matrix is considered for each class and how instances from that class receive various classifications. For class "b", 10 correctly classified instances were obtained but 6 was put into class "c", while for class "C", 29 were correctly classified but 4 were put into class "b".

To determine the attributes that have the most effect on the decision of patients staying at the hospital after surgery, the first model of 50 records is created using the rule induction algorithm, which is a PART algorithm. The results show that the rules that affect the patient's decision on post-surgery stay at a hospital are as follows:

1. COMFORT >7 AND BP-STBL = MOD-STABLE.
2. CORE-STBL = stable AND COMFORT >7 AND L-CORE = mid.

The attributes that affect their decision are as follows: COMFORT, BP-STBL, CORE-STBL, L-CORE.

Bayes networks

The Bayesian algorithm (Bayes' net) is used via WEKA on the post-operative patients' dataset of 50 records [21], [22]. In the present paper, the numbers of correctly and incorrectly classified instances are focused upon. Except for the Kappa statistic, the remaining statistics determine specific error quantities based on the class probabilities assigned by Bayes' net. Also, the confusion matrix is considered for each class and how instances from that class receive various classifications. For class "b", 6 correctly classified instances were obtained but 10 were put into class "c", while for class "C", 30 were correctly classified but 3 were put into class "b". To determine the exact attributes that most affect patients' decision to stay at the hospital after surgery, the first model of 50 records was used via WEKA using the Bayesian approach algorithm, which is Bayes' net algorithm.

The produced ADtree shows the main attributes that have the main effect on the decision of staying patients after surgery. The ADtree illustrates the probability of the attributes that affect the decision of patients staying after surgery. If we press on any attribute's circle, we get the probability distribution of a patient either staying, leaving, or being sent to the ICU, as illustrated in Figure 2.

decision	high	mid	low
I	0.2	0.2	0.6
S	0.2	0.429	0.371
A	0.188	0.565	0.246

decision	high	mid	low
I	0.2	0.6	0.2
S	0.2	0.429	0.371
A	0.275	0.449	0.275

decision	excellent	good	fair	poor
I	0.167	0.5	0.167	0.167
S	0.361	0.583	0.028	0.028
A	0.357	0.614	0.014	0.014

decision	high	mid	low
I	0.6	0.2	0.2
S	0.314	0.657	0.029
A	0.42	0.507	0.072

decision	stable	mod-stable	unstable
I	0.2	0.2	0.6
S	0.486	0.029	0.486
A	0.478	0.014	0.507

decision	stable	mod-stable	unstable
I	0.6	0.2	0.2
S	0.829	0.029	0.143
A	0.942	0.014	0.043

decision	stable	mod-stable	unstable
I	0.6	0.2	0.2
S	0.6	0.143	0.257
A	0.449	0.333	0.217

decision	'All'
I	1
S	1
A	1

Figure 2. Probability distribution of any of the attributes that affect the patient's decision to either stay, go home or being sent to the ICU.

In the above figure, the probability distribution of an attribute is obtained by pressing on any attribute's circle. These attributes affect the patient decision. In the present paper, the focus was only on the probability distribution of the attributes that affect the patient's decision. These attributes are as follows, based on the probability distribution table of each attribute showing the probability of a patient staying after surgery: L-CORE, L-SURF, L-O2, L-BP, SURF-STBL, CORE-STBL, BP-STBL, COMFORT), as illustrated above for each attribute.

ANALYSIS OF THE RESULTS

As mentioned above, four techniques were implemented to build classification models using 50 records of post-operative patients' datasets via WEKA. Table 2 illustrates the comparison between these four classification techniques.

Table 2. Analysis of the result of the implemented data mining classification techniques.

Number of Records	NN		RULES		BAYES		TREE		The percentage of staying patients	The percentage of out-patients	The total percentage
50	IN	OUT	IN	OUT	IN	OUT	IN	OUT	31%	9%	40%
	32	15	29	10	30	6	33	2			

The above table shows the results of implementing data mining classification techniques used to build models to predict the number of patients staying or leaving a hospital after surgery. Furthermore, it shows a comparison between four classification techniques and their algorithms for only the correctly classified records, which can be considered as the best and worst classification techniques, respectively. This can be determined by the error level of each classification technique and can also be considered as an advantage in data mining classification techniques.

In the present experiment, the best-performed classification technique is neural network since each implementation of its algorithm presents a higher rate of correctly classified records and fewer errors. Neural Networks are one of many data mining analytical tools that can be utilized to make predictions about key healthcare indicators such as cost or facility utilization. Neural networks are known to produce highly accurate results in practical applications. A typical neural network may have several hidden layers, and each layer can have several neurons. A neuron receives several signals from its input links, computes a new activation level, and sends it as an output signal through the output links. The input signal can be raw data or outputs of other neurons. The output signal can be either a final solution to the problem or ~~an~~ input to other neurons. The worst performed classification technique is the decision tree as its algorithms yield the highest error rate and lowest correctly classified records. Another advantage of using data mining classification techniques is in determining the exact attributes that have the most influence on a patient's decision. As illustrated above, the attributes were examined using the models created via WEKA.

DETERMINING BED REQUIREMENTS FOR POST-OPERATIVE PATIENTS

After determining the attributes that most affect the decision of a patient to stay or leave after surgery using data mining classification techniques, this section concentrates on post-surgery patients staying at a hospital. These patients need services such as beds, nurses and other services, which pertain to, and affect the management side of a hospital.

Chi-square test

A chi-square fitness test was used to evaluate the hypothesis that a random sample of size n of random variable X follows a particular distributional structure. This test formalizes the intuitive concept of comparing the data histogram to the shape of the candidate density or mass function [23]. The test is valid for large sample sizes, both for discreet and continuous distributional assumptions, where the parameters are determined by maximum probability. The test procedure starts with the arrangement of n observations into a set of K class intervals or cells, and the statistic is given by the following equation:

$$x^2 = \sum_{i=1}^{k} (O_i - E_i)^2 / E_i \quad (1)$$

where O_i and E_i are the observed and expected frequencies, respectively, for the i-th class interval. The expected frequency for each class interval is computed as $E_i = nP_i$, where P_i is the theoretical, hypothesized probability associated with the i-th class interval [24].

It can be shown that X^2 approximately follows the chi-square distribution with $K - S - 1$ degree of freedom, where S represents the number of parameters of the hypothesized distribution estimated by the sample statistics [24]. The hypotheses are as follows:

H_0: The random variable, X, conforms to the distributional assumption with the parameters (s) given by the parameter estimate (s); and

H_1: The random variable X does not conform.

In applying the test, if expected frequencies are too small, X^2 will reflect not only the departure of the observed frequency from the expected one but also the smallness of the expected frequency. Although there is no general agreement regarding the minimum size of E_i, values 3, 4, and 5 have been widely used. If the E_i value is too small, it can be combined with the expected frequencies in adjacent class intervals. The corresponding O_i values should also be combined and K should be reduced by one for each combined cell [25].

If the distribution being tested is discrete, each value of the random variable should be a class interval, unless it is necessary to combine adjacent class intervals to meet the minimum expected cell-frequency requirement. For the discrete case, if combining adjacent cells is not required,

$$P_i = P_{(x_i)} = P(X = x_i) \tag{2}$$

Otherwise, P_i is determined by summing the probabilities of appropriate adjacent cells.

PATIENTS STAYING AFTER SURGERY

Beds needed for each patient after surgery are decided using the queuing theory, which requires analyzing the number of patients staying after surgery and the number of days they will stay.

The Poisson distribution is named after the French mathematician Simeon Denis Poisson, who published its derivation in 1837 [26]. It is a discreet distribution of probabilities in which the likelihood of occurrence of an outcome with a short period of time is too small and the likelihood of two or more such outcomes within a short time is negligible. The occurrence of an outcome over a span of time is independent of each other. Poisson distribution's formula:

$$f(X) = e^{-\lambda}\lambda^x / x! \tag{3}$$

where,

λ: is an average rate of value in a given time of period.

x: is a Poisson random variable.

e: is the base of logarithm (e=2.718).

Table 3. Testing for the Poisson distribution.

x	freq	p	x*freq	p(x)	Ei	Ei	Oi	K
0	0	0	0	0.022736	0.841250443			
1	10	0.27027	10	0.08603	3.183109785			
2	8	0.216216	16	0.162759	6.022099594	10.04646	16	3.528072702
3	5	0.135135	15	0.205282	7.595440929	7.595441	5	0.88688908
4	3	0.081081	12	0.194186	7.184876554	7.184877	3	2.437507679
5	2	0.054054	10	0.146951	5.437203879	5.437204	2	2.172876127
6	2	0.054054	12	0.092672	3.428867311	6.730036	9	0.765633184
7	1	0.027027	7	0.050093	1.85344179			
8	2	0.054054	16	0.023693	0.876627874			
9	1	0.27027	9	0.009961	0.368552259			
10	1	0.27027	10	0.003769	0.139452206			
11	1	0.27027	11	0.001296	0.047968818			
12	1	0.27027	12	0.000409	0.015125303			

In the above table, the data was collected from 2–3 months of hospital records on the number of operations per day, the patient's admission and discharge dates. The data is randomly picked from the population, it is presented in the form of frequencies (observed and expected frequencies). The Chi-square test is implemented to test the number of patients staying after surgery. The results presented in Table 4 use the chi-square test, indicating that the data (The arrival time at a serviced location) indeed follows the Poisson distribution.

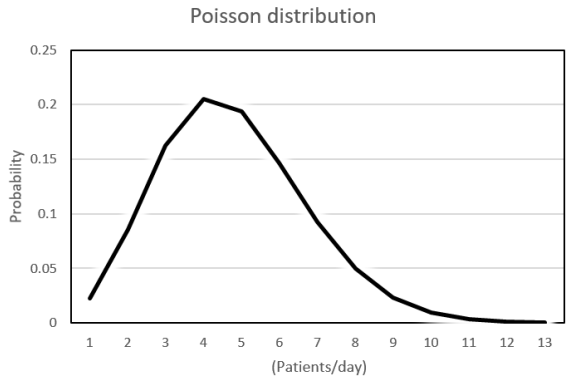

Figure 3. Poisson distribution.

The above figure illustrates the number of events (operations per day) following the implementation of the Poisson distribution formula. The curve represents values from the fit of the Poisson distribution given in columns 1 and 5 in Table 3. 2–3 months of hospital records on the number of operations, the patient's admission, and discharge dates from daily patients arrival were calculated as given in Table 3.

Duration of patients staying after surgery

Service time in the queuing model usually follows a negative exponential distribution. The data of the number of days a patient stays after surgery were tested for the negative exponential distribution. In probability theory and statistics, exponential distribution (also known as a negative exponential distribution) is a family of continuous probability distributions. It defines the time between events in the Poisson process, i.e. a process in which events take place continuously and independently at a constant average rate [27]. The Exponential distribution formula is defined by:

$$p(X) = \lambda e^{-\lambda x} \quad \text{for positive } \lambda \text{ and nonnegative } x \tag{4}$$

Table 4. Testing for the Exponential distribution.

X	freq	p	x*freq	p(x)	Ei	Ei	Oi	K
0	0	0	0	0.3034483	13.35172414	13.351724	0	13.35172414
1	17	0.386364	17	0.2240262	9.85715167	9.8571517	17	5.175966038
2	7	0.159091	14	0.1653914	7.277220382	7.2772204	7	0.010560507
3	5	0.113636	15	0.1221032	5.372539479	5.3725395	5	0.025832414
4	5	0.113636	20	0.0901449	3.966374376	6.8946219	7	0.001610608
5	2	0.045455	10	0.0665511	2.928247573			
6	2	0.045455	12	0.0491325	2.161831697	7.2723839	8	0.072799399
7	1	0.022727	7	0.036273	1.596011324	13.351724	0	13.35172414
8	1	0.022727	8	0.0267792	1.178284207			
9	1	0.022727	9	0.0197702	0.869889613			
10	1	0.022727	10	0.0145957	0.642211729			
11	1	0.022727	11	0.0107756	0.474124417			
12	1	0.022727	12	0.0079552	0.35003092			

In the above table, the data was collected from 2–3 months of hospital records on the number of operations per day, the admission as and discharge dates of patients. The data is randomly picked from the population, it is presented in the form of frequencies (observed and expected frequencies). The results presented in Table 4 using the chi-square test, indicating that the data (the number of days a patient stays after surgery) indeed follows an exponential distribution.

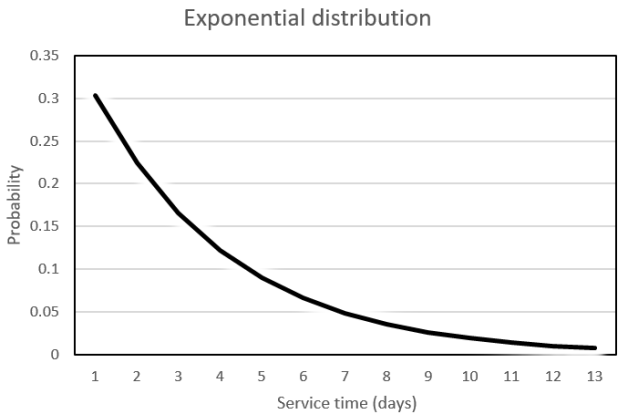

Figure 4. Exponential distribution.

The above figure illustrates the probability of a patient occupying a bed after surgery (bed service time, which is the waiting time for a patient to get a bed after surgery) following the implementation of the exponential distribution formula. The curve represents values from the fit of the exponential distribution given in columns 1 and 5 in Table 4. 2–3 months of hospital records on the number of operations per day, service time from daily patient's arrival were calculated as given in Table 3.

RESULTS

The results obtained from implementing both formulas were used by a simulation software (Queueing Theory Software for Calc**)** to obtain the number of beds required for each patient staying after surgery. Following this, a patient's waiting time in the queue after surgery is obtained. Queuing theory software for Calc provides a set of OpenOffice Calc spreadsheets that solve various queueing models. Collectively these spreadsheets are known as *QtsPlus4Calc*. The work of queueing software project is based on the Microsoft Excel- QtsPlus software package. The QtsPlus4Calc collection of spreadsheets will encompass the following areas: basic probability calculators; solvers for finite, linear difference equations; single-server, multiple-server, bulk, priority, and network analytic queueing models and several simulation models. Using a queuing model with exponential arrival and service time distributions, the following results were fed into the simulation software (Queueing Theory Software for Calc**)** to obtain the number of beds needed for patients staying after surgery:

- Arrival rate (λ =3.78 patient/day);
- Service rate (μ = 0.303951);
- Number of server (c = 13);

In the present paper, we considered an M/M/C queue in which the number of beds is fixed. The patients arrive according to a Poisson process at a rate (λ =3.78 patient/day). The service is provided by C servers, who serve the patients on a first-come-first-served (FCFS) basis. The service time of each patient is exponentially distributed with a mean ($1/\mu$=3.29 days). The results show that the probability of the customer arriving is delayed in a queue for about 0.826303, meaning that a patient's waiting time is approximately 1 with 13 instances and with 15 beds in the whole hospital. The hospital should thus provide a bed for a patient who is waiting in the queue by discharging recovered patients.

CONCLUSION

Data mining technologies can be of great value to the healthcare industry. However, healthcare data mining can be constrained by the availability of data, as raw data mining inputs frequently occur in various settings and structures, such as administration, hospitals, laboratories, and more. Data must also be obtained and implemented before data mining can be undertaken. In this paper, we used data mining classification techniques to build an analytical procedure based on raw data containing several attributes on patient status after surgery collected from the surgical department and the UCI Machine Learning Repository. These data mining models helped determine which patients will stay and which will go home after surgery and also the attributes that contribute to the patient's decision to stay. Further, the present work indicates which classification technique is the most appropriate out of the four techniques (neural network, decision tree, rule induction, and Bayesian network). Also, the study helps determine which out of all the nine attributes have the most effect on the decision of a patient staying or leaving after surgery. It was proved that neural networks performed better than the other DM techniques in terms of accuracy and prediction. After determining the attributes affecting the stay of a patient, the next task was to determine the number of adequate beds required to accommodate them by analyzing the duration of the stay. The queuing model used by the Queueing Theory software allowed us to determine the number of beds that will keep the wait length to 1. It is assumed that if the length is 1, then one of the current patients occupying a room will be discharged, which is a common hospital policy. The study can further help any healthcare organization benefit from the developed data mining application in the management aspect of health care and could potentially contribute to the research on data mining and the healthcare industry. Future work will investigate other areas where data

mining techniques can be used in the healthcare industry to help with the computerization of business and decision-making processes.

REFERENCES

[1] I. Witten, E. Frank, and M. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Cambridge, USA : Morgan Kaufmann, pp, 2016, p. 655.

[2] M. Hammer and J. Champy, *Reengineering the Corporation: Manifesto for Business Revolution*. London: Brealey Pub, 2009, p. 272.

[3] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Amsterdam: Elsevier/Morgan Kaufmann, 2012, p. 744.

[4] S. C. Pandey, "Data mining techniques for medical data: A review," *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, 2016, pp. 972-982.

[5] W. Fan and A. Bifet. "Mining big data," *ACM SIGKDD Explorations Newsletter*, vol. 14, no. 2, pp. 1–5, 2013.

[6] M. Silver, T. Sakata, H. C. Su, C. Herman, M. J. O'Shea, and S. B. Dolins, "Case study: How to apply data mining techniques in a healthcare data warehouse," *J. Healthc. Manag.*, vol. 15, no. 2, pp. 155–164, 2001.

[7] N. Jothis, N. A. A. Rashid, and W. Husain, "Data mining in healthcare – A review," *Procedia Comput. Sci.*, vol. 72, pp. 306–313, 2015.

[8] M. J. Zaki and W. Meira, *Data Mining and Analysis: Fundamental Concepts and Algorithms*, New York, NY, USA: Cambridge Univ. Press, 2017.

[9] O. Folorunso and A. Ogunde, "Data mining as a technique for knowledge management in business process redesign," *Information Management & Computer Security*, vol. 13, no. 4, pp.274–280, 2005.

[10] M. Islam, M. Hasan, X. Wang, H. Germack, and M. Noor-E-Alam, "A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining," *Healthcare*, vol. 6, no. 2, p. 54, May 2018.

[11] M. Persson and J. Persson, "Analysing management policies for operating room planning using simulation", *Health Care Management Science*, vol. 13, no. 2, pp. 182-191, 2010.

[12] K. Kasikumar, M. MohamedNajumuddeen and R. Suresh, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks." *International Journal of Data Mining Techniques and Applications* vol 7, pp.172-176, 2018.

[13] M. Islam, M. Hasan, X. Wang, H. Germack and M. Noor-E-Alam, "A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining", *Healthcare*, vol. 6, no. 2, p. 54, 2018.

[14] P. Ahmad, S. Qamar and S. Qasim Afser Rizvi, "Techniques of Data Mining In Healthcare: A Review", *International Journal of Computer Applications*, vol. 120, no. 15, pp. 38-50, 2015.

[15] P. Murphy and D. Aha, "UCI Machine Learning Repository", 1992. [Online]. Available: http://archive.ics.uci.edu/ml. [Accessed: 23- Aug- 2020].

[16] J. Bell, *Machine Learning: Hands-On For Developers And Technical Professionals*, 2nd ed. Wily, 2020.

[17] J. David, "Data Mining and Data Warehousing", *The SIJ Transactions on Computer Science Engineering & its Applications (CSEA)*, pp. 17-19, 2019.

[18] A. Segatori, F. Marcelloni and W. Pedrycz, "On Distributed Fuzzy Decision Trees for Big Data," in IEEE Transactions on Fuzzy Systems, vol. 26, no. 1, pp. 174-192, Feb. 2018.

[19] A. I. Georgevici and M. Terblanche, "Neural networks and deep learning: a brief introduction," *Intensive Care Medicine*, vol. 45, no. 5, pp. 712-714, Feb. 2019.

[20] O. Akgobek, "A Rule Induction Algorithm for Knowledge Discovery and Classification," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 21, no. 5, pp. 1223-1241, 2013.

[21] J. Cheng and R. Greiner, "Comparing Bayesian Network Classifiers." *UAI'99: Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, Alberta, 1999, pp. 101–108.

[22] R. Pi, Y. Cai, Y. Li and Y. Cao, "Machine Learning Based on Bayes Networks to Predict the Cascading Failure Propagation," in *IEEE Access*, vol. 6, pp. 44815-44823, 2018.

[23] S. Kilic, "Chi-square Test", *Journal of Mood Disorders*, vol. 6, no. 3, p. 180, 2016.

[24] A. Richardson, "Nonparametric Statistics: A Step-by-Step Approach", *International Statistical Review*, vol. 83, no. 1, pp. 163-164, 2015.

[25] G. Zacharewicz, "Generalized Discrete Event System specification: A state-of-the-art study", *Simulation*, 2018.

[26] E. Cepeda, "Stochastic coalescence multi-fragmentation processes", *Stochastic Processes and their Applications*, vol. 126, no. 2, pp. 360-391, 2016.

[27] K. Balakrishnan, Exponential Distribution: Theory, Methods and Applications, London: Routledge, 2020.