

Development Hospital Admission System Using Data Mining Tasks With Security

¹ Panta Nrushimha Saraswathi, ² Dr.P.Pedda Sadhu Naik

Abstract: Healthcare organizations often benefit from information technologies as well as embedded decision support systems, which improve the quality of services and help preventing complications and adverse events. Owing to the great advantages various organizations are using data mining technology is a vital part for everyone. We propose a model for predicting the patient length of stay (LOS) in ED using data mining techniques. The used data was collected from the pediatric emergency department (PED) in Lille regional hospital centre. Therefore it gets important to find innovative methods to improve patient flow and prevent overcrowding. We identified two models based on linear regression. These models are validated and were successfully applied to the classification and prediction of the LOS in the pediatric emergency department (PED) at Lille regional hospital centre, France. It is shown how the health care industry is solving their problems through data mining techniques. Various learning methods in data mining, data mining tasks, importance of data mining in health care industry, forecasts and issues in health care industry are discussed. It focus on developing and applying machine learning and data mining tools to an array of different challenging problems from clinical genomic analysis, through designing clinical decision support systems.

Index Terms: Data Mining, Health care, Clustering, Classification, Extraction, Predictions, Emergency department, healthcare modeling, Machine Learning, prediction, length of stay.

1. INTRODUCTION

Today, Health organizations are capable of generating and collecting a large amount of data. This increase in data volume automatically requires the data to be retrieved when needed. With the use of data mining techniques is possible to extract the knowledge and determine interesting and useful patterns [1]. The objective is to propose a Decision Support System (DSS) based on a data mining approaches, in order to support operational and tactical decisions [2]. According to the patient symptoms the patient is categorized into three levels of triage process. In this, critical patients are given first importance and non-critical patients are given a waiting time according to their symptoms to get treated by doctors [3]. The aim of this study is to see how can we use data mining techniques and particularly classification methods to develop models for prediction of patients' length of stay (LOS) at the emergency department [4]. This paper aims to provide an overview of the data mining starting from its definition, to the application, how this technique works together with a summary some of the recent work in the healthcare sector [5]. Knowledge driven approaches operate on knowledge repositories that include scientific literature, published clinical trial results, medical journals, textbooks, as well as clinical practice guidelines [6]. There are two general categories of algorithms: unsupervised and supervised. Unsupervised machine learning

algorithms are typically used to group large amounts of data [7]. Unsupervised algorithms can be used to generate hypotheses, and thus often precede use of a supervised algorithm. Supervised machine-learning algorithms start out with a hypothesis and categories that are set out in advance [8].

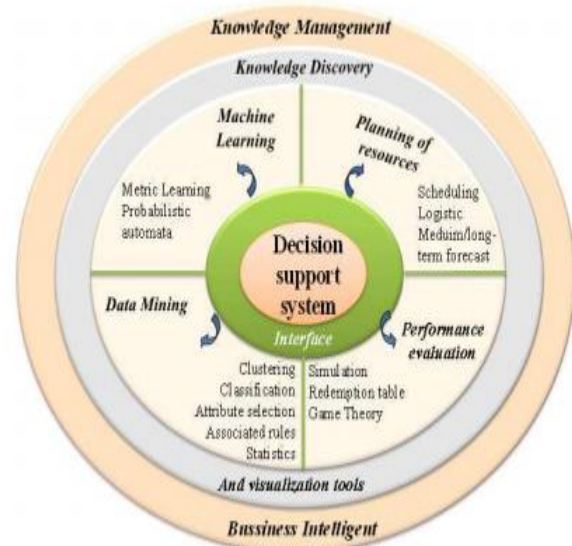


Figure 1: Modelling environment.

2. RELATED WORK

Data mining allows finding models and patterns from the available data. Data mining is a discipline at the interface of statistics and information technologies:

databases, artificial intelligence and machine learning [9]. In contrast for supervised learning a model is built prior to the analysis. We then apply the algorithm to the data in order to estimate the parameters of the model. The objective of building models using supervised learning is to predict an outcome or category of interest [10]. The model uses regression which is actually a time series regression and exponential smoothing where time series regression performs better than linear regression [11]. Used artificial neural network ensembles to predict the disposition and length of stay in children presenting to the Emergency Department with bronchiolitis. The results show that the neural network ensembles correctly predicted disposition in 81% of test cases [12]. This algorithm is compared with various classification algorithms such as Neural Network, Logistic Regression, Random Forest and Decision Tree, SVM, were used to compare the proposed algorithm. Different parameters such as Precision, Accuracy, Recall and Area under the Curve (AUC) were to measure the performance of the proposed algorithm [13]. One study conducted to determine the factors affecting LOS in public hospitals in Loreto Province, Iran demonstrated that, first, an increase in age would lead to an increase in average LOS and, second, the average LOS of men is longer than that of women [14].

3. SYSTEM ARCHITECTURE

The pediatric emergency department (PED) is open internal capacity, the PED shares many resources, such as administrative patient registration and clinical laboratories, with other hospital departments. In this study the data mining based approach is presented and applied to predict the patient's length of stay at the PED [15]. It uses a combination of mathematical and computational techniques to aid description and classification, and to extract knowledge of data set [16]. They are reliable and have better accuracy in clinical decision-making decision trees are the most current decision tree algorithms. The next algorithm is the decision tree which is specifically recursive partitioning. The algorithm splits the data at each node based on the variable that separates the data unless an optimal model is not obtained [17].

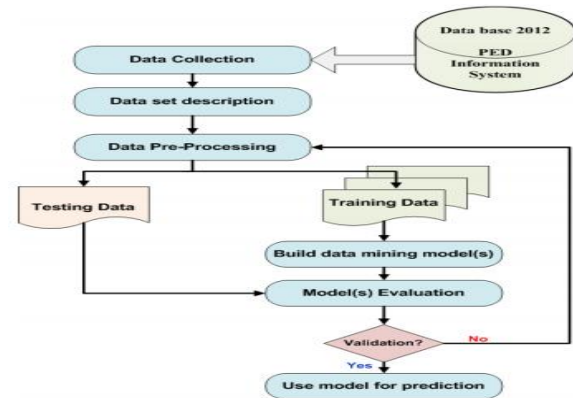


Figure 2: General architecture of the proposed approach

4. METHODOLOGY

The proposed methodological framework is based on predictive modeling to approximate the LOS of a new patient at the ED. The goal is to enable prediction with a very useful model, easily understandable and usable by the hospital staff [18]. There is a wealth of knowledge to be gained from computerized health records. Yet the overwhelming bulk of data stored in these databases makes it extremely difficult, if not impossible, for humans to sift through it and discover knowledge [19]. To obtain reliable results, the extracted knowledge should be evaluated by a comparison of results obtained with various supervised classification methods and using several measures.

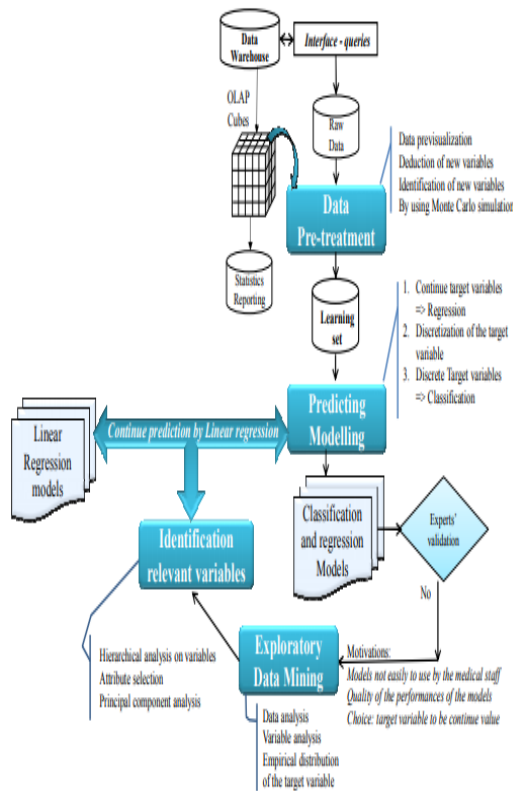


Figure 3: Proposed methodology

A. Classification Algorithm

Although many interesting real world domains demand for regression tools, machine learning researchers have traditionally concentrated their efforts on classification problems. Classification algorithms only deal with nominal variable and cannot handle ones measured on a numeric scale. To use them numeric attributes must first be “discretized” into smaller number of distinct ranges [20]. We decide to take into account the intuitive staff’s experience in order to estimate the intervals more realistic than the values obtained by Weak tools for dies cartelization.

1. Logistic Model Tree (LMT)
2. Multi-class alternating decision tree using the Logit Boost strategy (LAD Tree)
3. Decision tree (C4.5 - J48),
4. Decision tree with naive Bays classifiers at the leaves (NB Tree)
5. Random Forest (RF),

6. Decision/regression tree using information gain/variance and prunes it using reduced-error pruning (REP Tree)
7. Multilayer Perception (MP),
8. SVM (SMO).

For these eight methods, the first step corresponds the target variable LOS in order to apply some of these methods. The model valuation is carried out using 10-fold cross validation on the PED data set [21].

5. DATA MINING TASK

The data mining tasks are different types depending on the use of data mining result the data mining tasks are classified [22].

A. Exploratory Data Analysis In the repositories vast amount of information’s are available .This data mining task will serve the two purposes (i)Without the knowledge for what the customer is searching, then (ii) It analyze the data these techniques are interactive and visual to the customer.

B. Descriptive Modeling It describe all the data, it includes models for overall probability distribution of the data, partitioning of the dimensional space into groups and models describing the relationships between the variables.

C. Predictive Modeling This model permits the value of one variable to be predicted from the known values of other variables.

D. Discovering Patterns and Rules. This task is primarily used to find the hidden pattern as well as to discover the pattern in the cluster. In a cluster a number of patterns of different size and clusters are available ..

E. Retrieval by Content The primary objective of this task is to find the data sets of frequently used in the for audio/video as well as images It is finding pattern similar to the pattern of interest in the data set.

5.1 Techniques for Continue Target Value

Only regression allows us to use continue variable of the target variable. Nine methods available are used [23].:

1. **LR:** Linear regression
2. **DS (Decision Stump):** Class for building and using a decision stump. Usually used in conjunction with a boosting algorithm regression classification
3. **MSP:** Induction of model trees for predicting continuous classes

4. REP Tree: Builds a decision/regression tree using information gain/variance and prunes it using reduced-error pruning (with backfitting).

5. SM Oreg: SVM Regression

6. IRM: Isotonic regression model

7. MLP: MultiLayer perceptron

8. PRLM: Pace regression linear models

9. K Star: K-nearest neighbor's classifier (lazy)

Classification models are not very useful when the target values are continuous. We also observe that independently of the used algorithms, we obtain quasi the same performances.

6. RESULTS AND DISCUSSION

For the evaluation of the methods accuracy, kappa, sensitivity and specificity this performance metrics are used. As shown in table, the Random forest performs best across all performance measures. A small difference is observed the remaining two methods decision tree and gradient boosted machine. For admission of any patient through the emergency department the triage process plays an important role. It is based on the partition of the original sample into ten subsamples, using nine as training data and one for testing data. As shown by results, this study can be benefit to PED manager to predict LOS and detect the beginning of strain situation. This information can be used to make the decision more proactive.

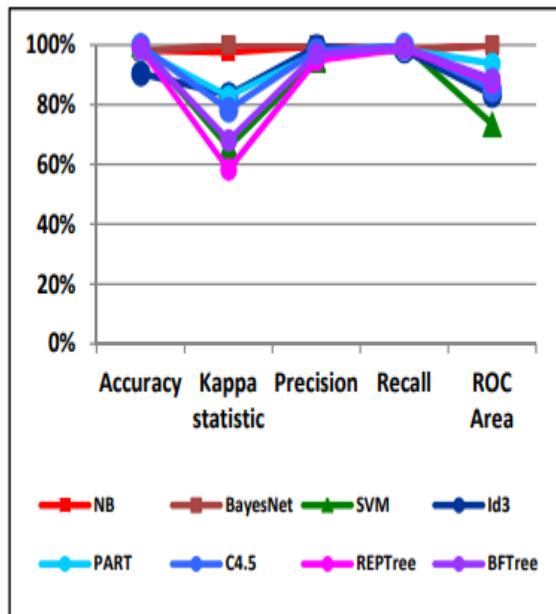


Figure 4: Rate of maximum

7. CONCLUSION AND FUTURE WORK

The overall study involved a survey of different methods used for the prediction model of hospital admission. We showed that different data mining techniques can be used beneficially in classification and prediction by using linear regression and we obtain a very “simple” model that it can easily use by the medical staff in order to estimate the LOS. The different methods of data mining are used to extract the patterns and thus the knowledge from this variety databases. Selection of data and methods for data mining is an important task in this process and needs the knowledge of the domain. This study illustrates the benefit of classification methods to make PED management system more proactive face to strain situations. The benefit not only include prediction of medical condition using the previous history of a patient from the database but also hospital management systems such as emergency division. In future, different algorithms regarding deep learning and machine learning can used to implement the model. Even ensemble of different algorithms can also be done. Different demographics as predictor can be taken into consideration.

8. REFERENCES

- [1] J.S. Olshaker, N.K. Rathlev, Emergency Department overcrowding and ambulance diversion: The impact and potential solutions of extended boarding of admitted patients in the Emergency Department, *J. Emerg. Med.* 30 (2006) 351–356.
- [2] Gomez V, Abasolo JE. Using data mining to describe long hospital stays. *Paradigma.* 2009; 3(1):1–10.
- [3]. 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):27–38.
- [4]. Chang KC, Tseng MC, Weng HH, Lin YH, Liou CW, Tan TY. Prediction of length of stay of first-ever ischemic stroke. *Stroke.* 2002; 33(11):2670–2674.
- [5]. Jiang X, Qu X, Davis L. Using data mining to analyze patient discharge data for an urban hospital. In : *Proceedings of the 2010 International Conference on Data Mining*; 2010 Jul 12-15; Las Vegas, NV. p. 139–144.
- [6]. Isken MW, Rajagopalan B. Data mining to support simulation modeling of patient flow in hospitals. *J Med Syst.* 2002; 26(2):179–197.

- [7]. Walczak S, Scorpio RJ, Pofahl WE. In : Cook DJ, editor. Predicting hospital length of stay with neural networks. Proceedings of the Eleventh International FLAIRS Conference; 1998 May 18-20; Sanibel Island, FL. Menlo Park, CA: AAAI Press;1998. p. 333–337.
- [8]. 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):211–221
- [9] S. L. Nalawade and R. V. Kulkarni, "Application of Data Mining in Health Care," *International Journal of Science and Research (IJSR)*, vol. 5, no. 4, pp. 262-268, 2016.
- [10] P. Chantamit-o-pas and M. Goyal, "Prediction of Stroke Using Deep Learning Model" D. Liu et al. (Eds.): *ICONIP, Part V, LNCS 10638*, pp. 774–781, 2017.
- [11] D. Tomar and S. Agarwal, "A Survey on Data Mining Approaches for Healthcare," *International Journal of Bio-Science and BioTechnology*, vol. 5, no. 5, pp. 241- 266, 2013.
- [12] N. P. Waghulde, N. P. Patil, "Genetic Neural Approach for Heart Disease Prediction," *International Journal of Advanced Computer Research*, vol. 4, no. 3, pp. 778-784, 2014
- [13] Zhang, Xingyu and Kim, Joyce and Patzer, Rachel E and Pitts, Stephen R and Patzer, Aaron and Schrage, Justin D, "Prediction of emergency department hospital admission based on natural language processing and neural networks," *Methods of information in medicine.*, vol.56, pp. 377– 389, 2017.
- [14] Boukenze, Basma and Mousannif, Hajar and Haqiq, Abdelkrim, "Predictive analytics in healthcare system using data mining techniques," *Computer Sci Inf Technol.*, vol. 1, pp. 1–9, 2016.
- [15] Dinh et al., "The Sydney Triage to Admission Risk Tool (START) to predict Emergency Department Disposition: A derivation and internal validation study using retrospective state-wide data from New South Wales, Australia," *BMC emergency medicine.*, vol.16, pp. 46, 2016.
- [16] Golmohammadi, Davood, "Predicting hospital admissions to reduce emergency department boarding," *International Journal of Production Economics.*, vol. 182, pp. 535–544, 2016
- [17]. J. W. Lee, J. B. Lee, M. Park, and S. H. Song, "An extensive comparison of recent classification tools applied to microarray data," *Computational Statistics & Data Analysis*, Vol. 48, No. 4, pp. 869–885, 2005.
- [18]. Met-Office. The Met Office health forecasting services, Exeter. 2009.
- [19]. Boyle J, Jessup M, Crilly J, Green D, Lind J, Wallis M, et al. Predicting emergency department admissions. *Emerg Med J.* 2011
- [20]. Gentry L, Calantone RJ, Cui SA. The forecasting classification grid: a typology for method selection. *J Glob Bus Manag.* 2006; 2(1):48–60.
- [21]. Kamath C. *Scientific data mining: a practical perspective*. Philadelphia (PA): Society for Industrial and Applied Mathematics;2009.
- [22]. Son YJ, Kim HG, Kim EH, Choi S, Lee SK. Application of support vector machine for prediction of medication adherence in heart failure patients. *Healthc Inform Res.* 2010; 16(4):253–259.
- [23]. Kajabadi A, Saraee MH, Asgari S. Data mining cardiovascular risk factors. In : *Proceedings of International Conference on Application of Information and Communication Technologies*; 2009 Oct 14-16; Baku, Azerbaijan. p. 1–5.

Authors Details:



PANTA NRUSHIMHA SARASWATHI EMAIL : nrushimhasaraswathi@gmail.com
Dr.Samuel George Institute of Technology,
Markapur, AP.

Dr.P.Pedda Sadhu Naik, Professor & HOD,
Dr.Samuel George Institute of Technology,
Markapur, AP.