ISSN 2347 - 3983
Volume 8. No. 5, May 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter84852020.pdf
https://doi.org/10.30534/ijeter/2020/84852020

Hospital Bed Support System Based on Machine Learning

Dung Q. Tran[1], Lien T. Tran[2], Binh A. Nguyen[3], Viet Q. Tran[3], Nhan D. Nguyen[4], and Giao N. Pham[5]
[1]Advances Analytics & AI Center, FPT Software, QuyNhon, Vietnam, dungtq20@fsoft.com.vn
[2]Dept. of Information Technology, QuyNhon University, BinhDinh, Vietnam, tranthilien@qnu.edu.vn
[3]ICT Department, FPT University, Hanoi, Vietnam, binhnase04865@fpt.edu.vn, viettqse06178@fpt.edu.vn
[4]Dept. of Biomedical Engineering, Sungkyunkwan University, Suwon, South Korea, nhannd@skku.edu
[5]Dept. of Computing Fundamentals, FPT University, Hanoi, Vietnam, giaopn@fe.edu.vn

ABSTRACT

Recent years, the phenomenon of hospital overcrowding has become more and more severe and occurs in all levels with 2-3 people per bed, which has become an urgent issue for the health system as well as the whole society. One of the solutions to reduce this situation is to arrange beds in a reasonable way. Based on the large amount of data on examination and treatment in hospitals, we propose a solution to use Machine Learning to solve the above problem. A district hospital in Binh Dinh has provided 15066 medical records of 20 different hospitalizations, including information on patients such as year of birth, ethnicity, hometown, diagnosis, date admission and discharge date. With these records, we have built into a dataset and conducted experiments that predict the number of days the patient will be inpatient. The best experimental result is the use of XGBoost Regression, the R2 coefficient is about 0.84.

Key words: Machine Learning, Random Forest Regression, XG-Boost Regression, Linear Regression, Logistic Regression

1. INTRODUCTION

Currently, Machine Learning is being researched and applied in many different fields around the world. In Vietnam, this is still relatively new but it has also been studied and applied in a number of areas that bring benefits to society. Over the years, with the development of information technology and its application in many areas of social life, the amount of data collected and stored by hospitals has increased [1], [2]. This data is stored because it is thought to contain certain values as shown in figure 1. However, statistically, only a small amount of this data is analyzed and used by humans to benefit society [3], [10]-[12].

ID	Năm Sinh	Dân Tộc	Tên Quận Huyện	Tên Tỉnh Thành	Mã ICD	Chẩn Đoán	Ngày Vào	Ngày Ra
1.60E+17	2013	Kinh	Thanh pho Qui Nhon	Binh Dinh	B34	Sot sieu vi	12/28/2015 7:27	1/2/2016 7:00

Figure 1: Patient data example.

Nowadays, we go to the clinic after the diagnosis, the doctor will inform the time of inpatient treatment with the disease we encounter. The time given by the doctor will be based on the treatment regimen corresponding to that disease. However, we found that the inpatient treatment time of the same people is different because each person has a different condition, health.

In this paper, we used the regression algorithm to build a system to make predictions about inpatient treatment time appropriate to the body, health for different people. The proposed solution will be presented in Sec. 2. Sec. 3 describes experimental results, and the conclusion will be explained in Sec. 4.

2. THE PROPOSED SOLUTION

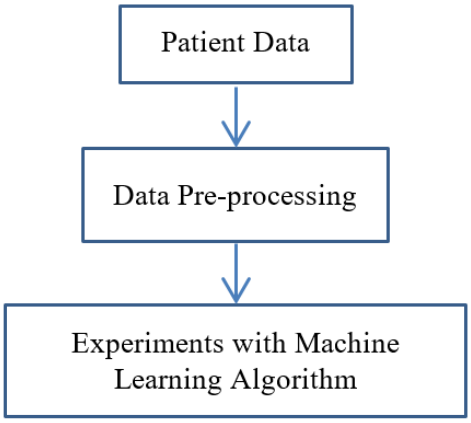

Figure 2: The proposed solution.

The proposed solution is shown in figure 2. The data we have is actual data, stored by the hospital in the hope that we can extract useful knowledge to better serve medical examination and treatment. With this data set, we want to build a system that makes predictions about the number of days a patient will be inpatient to help the hospital arrange a hospital bed. We need preprocessing because this data set also contains error values and formats that are not compatible with Machine Learning algorithms. In addition, we found that it was possible to infer new attributes from the original attribute such as the year of birth and the date of admission to the patient's age, the date of admission and the date of discharge will have the duration of treatment. In particular, we recognize that

different hospitalizations will have different safety values. Therefore, we will use the average value in combination with the standard deviation of each hospital admission to determine a threshold for a safe area.

The machine learning algorithm only accepts input as numeric values, so we need to transform numerical values with the smallest change in meaning of values. We will then experiment the data set with regression algorithms, because the treatment time value is continuous [4]. The algorithms we experiment with include: Linear Regression [5], Random Forest Regression [6], Logistic Regression [7], XG-Boost Regression [8].

Parameter R^2 [9] will be used to evaluate the experimental results of algorithms. It is computed by Eq. (1).

$$R^2 = 1 - \frac{ESS}{TSS} \quad (1)$$

There in ESS is Residual Sum of Squares, and TSS is Total Sum of Squares. The ESS and TSS are computed by Eq. (2) and Eq. (3) respectively. From Eq. (1) we could see that the value of R^2 always smaller than 0 and less than 1. The value of R^2will be equal 1 if the value of ESS is equal 0.

$$ESS = \sum_i (y_i - \hat{y}_i)^2 \quad (2)$$

$$TSS = \sum_i (y_i - \bar{y}_i)^2 \ with\ \bar{y}_i = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (3)$$

3. EXPERIMENTAL RESULTS

Table 1: Data from hospital.

Name	Number
Pneumonia	770
Cataracts of the nucleus in the elderly	485
Fever with no known cause	396
Shallow lesions affect many body areas	457
Thai paper	355
Enteritis	253
Fracture of the forearm	289
Cerebral infarction	280
Kidney stones	275
Inguinal hernia	373
Open wounds of the eyelids and the area around the eyes	271
Collarbone fracture	260
Diarrhea level	253
Digestive disorders	245
Caesarean section for a fetus	3343
Abortion is a pregnancy	2862
Dengue hemorrhagic fever	1218
Fever virus	947
Acute appendicitis	868
Hemodialysis	866

Table 2: Experimental Results.

Method	Accuracy (%)
Logistic Regression	54
Random Forest Regression	77
Linear Regression	81
XG-Boost Regression	84

Figure 3: Data after pre-processing step.

The data set we used consists of 15066 medical records of 20 different hospitalizations with the number shown in table 1. We use this data set to solve the problem of making an inpatient treatment time estimate suitable for each person with the same hospital admission. The data consists of 9 attributes which are the information of the patients including: ID code, year of birth, ethnicity, name of the district where they live, name of province where they live, ICD code (disease code), diagnosis of doctor, admission and discharge date.

We undergo a number of important preprocessing steps such as eliminating noise, error values, creating additional properties, transforming data and one-hot encoding. Some of the changes in the dataset are summarized as follows: ethnic group (Kinh, ethnic group), residence (rural, district, city), age (children, youth, middle-aged, elderly) as shown in figure 3. The data is divided into a ratio of 7-3, of which 70% is used for training and 30% of the evaluation. We tested this data set with regression algorithms including: Random Forest Regression, Linear Regression, Logistic Regression, XG-Boost Regression. Experimental results are summarized in table 2, the XG-Boost Regression algorithm has the best results when the R2 coefficient is 0.84.

4. CONCLUSION

We analyzed the collected data relatively well to come up with suitable data preprocessing options. With this data set, we have extracted useful knowledge to help the hospital take the initiative in arranging hospital beds, inpatient time and planning to buy drugs in time. In addition, knowledge can assist patients to arrange time and money for treatment. The experience gained from data analysis and pre-processing of data, we will continue to collect more data and processing to increase the accuracy of the model. In addition, we will collect and process data from other hospital admissions to increase the number of cases the model supports to assist in arranging hospital beds.

ACKNOWLEDGEMENT

This work is supported by FPT Software Company Co., Ltd., Hanoi, Vietnam; FPT University, Hanoi, Vietnam; Quy Nhon University, BinhDinh, Vietnam; and Sungkyunkwan University, Suwon, Republic of Korea.

REFERENCES

1. W. L. Klaus, P. W. Jonathan, and J. M. Clark. **Development and Validation of a Model for Predicting Inpatient Hospitalization,** Medical Care, Vol. 50, No. 2, pp. 131-139, Feb. 2012. https://doi.org/10.1097/MLR.0b013e3182353ceb
2. R. A. Taylor, J. R. Pare, A. K Venkatesh, H. Mowafi, E. R. Melnick, W. Fleischman, and M. K. Hall.**Prediction of In-hospital Mortality in Emergency Department Patients with Sepsis: A Local Big Data-Driven, Machine Learning Approach**, Acad. Emerg. Med. Vol. 23, No. 3, pp. 269-278, March, 2016. https://doi.org/10.1111/acem.12876
3. M. C. Kneyber, K. G. Moons, and D. R. Groot. **Prediction of duration of hospitalization in respiratory syncytial virus infection**, PEDIATRIC PULMONOLOGY, Vol. 33, No. 6, pp. 453-457, June 2002. https://doi.org/10.1002/ppul.10099
4. L. Andy, andW. Matthew. **Classification and Regression by Random Forest**, R News, Vol. 2, No. 3, pp. 18-23, 2002.
5. H. Z. Kelly,T. Kemal, and G. S. Stuart. **Correlation and Simple Linear Regression**, Mathematics, 2003.
6. M. R. Segal. **Machine Learning Benchmarks and Random Forest Regression**, Division of Biostatistics, University of California, San Francisco, CA 94143-0560, April, 2003
7. R. E. Wright. **Logistic Regression**, Reading and Understanding Multivariate Statistics, pp. 217-244, American Psychological Association, 2003.
8. T. Chen, andG. Carlos. **XG-Boost: A Scalable Tree Boosting System**, in Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794, 2016. https://doi.org/10.1145/2939672.2939785
9. A. C. Cameron,and A. G. W. Frank.**An R-Squared Measure of Goodness of Fit for Some Common Nonlinear Regression Models**, Journal of Econometrics, Vol. 77, No. 2, pp. 329-342, April 1997. https://doi.org/10.1016/S0304-4076(96)01818-0
10. B. Poggel, and B. Engel. **A Linear-Elastic Linear-Plastic Approach for the Forming Simulation of Sandwich Sheets with Cohesive**, International Journal of Emerging Trends in Engineering Research, Vol. 5, No. 6, pp. 11-16, June 2017.
11. A. S. Bassam, and A. Mohammed. **Performance of RPL in Healthcare Wireless Sensor Network**, International Journal of Emerging Trends in Engineering Research, Vol. 8, No. 3, pp. 797-803, March 2020. https://doi.org/10.30534/ijeter/2020/31832020
12. B. Abbas, and H. Ahmed. **Novel X-rays attenuation by (PMMA-PS-WC) New Nano-Compsites: Fabrication, Structural, Optical Characterizations and X-Ray Shielding Application**, International Journal of Emerging Trends in Engineering Research, Vol. 7, No. 8, pp. 131-144, August 2019.