



## Analysis of Data Mining Techniques for Prediction of Admissions in Hospital Emergency Department.

**Garima Sharma, Prof Krutika Chhajed**

*Master of Engineering First year Computer Science and Engineering Department, P.R.Pote (Patil) Education & Welfare Trust's Group of Institution, College of Engineering & Management, Amravati, (India)*

*Professor Computer Science And Engineering Department, P.R.Pote (Patil) Education & Welfare Trust's Group Of Institution, College of Engineering & Management, Amravati,(India)*

### **ABSTRACT**

*The patients may experience inconvenience whenever there is crowding in Emergency Departments (EDs). The quality of health care is affected due to crowding of ED. To avoid the adverse effects of inconvenience and improve patient care, there is a need to explore ideas and innovative technology methods for predictions of admissions in ED. The analysis of data gathered from hospitals like the patient age, previous history, month of the year, day of the week and time of the day in which the patient was admitted in the ED for health care work as the key for predictions of future admissions using data mining techniques with the help of some machine learning algorithms. The usage of data mining techniques combined with the following three techniques logistic regression, decision trees, and (GBM) Gradient Boosted Machines give the final method for predictions of patient admission in ED. The advantages of using data mining techniques for prediction are, hospitals can be more resourceful in managing health care. The combination of logistic regression method and decision tree models give an interpretation of patients in ED. The GBM model gives the accurate value, hence used when accuracy is the priority.*

**Keywords** *Data Mining, Decision trees, Emergency Department, Gradient Boosted Machines, logistic regression, Machine learning.*

### **1. INTRODUCTION**

The law passed by Emergency Medical Treatment and Active Labour Act (EMTALA) of 1986 stated that any person seeking medical attention should be provided with a complete medical screening examination regardless of the nationality, legal status and ability to pay the bill, hence it is necessary for Emergency Departments (EDs) in hospitals to give required attention to every patient visiting the hospital for medical care. If the staff present at the hospital is not sufficient for the health care of increasing number of incoming patients and the inpatients, the other hospitals are at a long distance from the patients, the capacity of beds available in the hospital is not enough for admitting all the ED patients and they have to be transferred to other nearby



hospitals for further treatment. In all the mentioned and many other possible cases the patients suffer more than any staff member of the hospital. Hence, the hospital acquired solutions from the new emerging technical methods.

One of the most efficient methods used by hospital EDs is Data Mining with some machine learning techniques. The previous data from the records of the hospital's EDs play an important role for extracting patterns. The model created by use of data mining techniques is helpful to increase the performance of ED. Stress caused to the waiting patients can be reduced by the previous prediction by keeping alert of the probable number of bed for the patients to be admitted, the storage of resources and all the necessary requirements of the EDs required for complete patient care. The result of the model designed by using data mining techniques, logistic regression model, gradient boosting machine can differ from the actual number. Prior to start the implementation of prediction model testing is important.

Analysis of the methods used in past, technique used in current system and the possibility of changes that can be done in previous and current model or looking for the requirement of a new perspective and ideas for enhancement is necessary. Individual concept and model were used in previous prediction systems. Combination of data pattern from data mining along with machine learning techniques requires to be tested and evaluated for changes and accuracy. The methods have their own individual properties which are seen in the output. The purpose of analysis is to find a model which is suitable for hospitals emergency department management to provide care to the patients without arising any bottleneck in the process.

## 2. RELATED WORKS

The analysis and design is not based on a hypothesis or any random guess by a person or a group. The solution cannot be given by simple guess work, the systematic approach to science and technical models are needed. The technical model designing require study of past data in the hospitals, each patient have a different health record hence analysis is done to identify some particular patterns in them.[3]The patients record can have pattern based on the record of heart rate, emergency severity index, triage score, blood pressure and major problem. For the convenience of the elderly patients in emergency department LaMantia and other members [2] found a logistic regression model to identify the pattern and probability of patients revisiting the hospital. The model was also to predict the accuracy for attendance of the hospital's staff present in ED.

Using historical data the predicting models were designed, Boyle and team [4] found the errors in their models by using mean absolute percentage error (MAPE). Sun and team [5] as well as Cameron and team [6] both the teams designed the models using logistic regression technique. They took into consideration the age, chronic conditions, week day, out of hour attendances in designing of their two different model for prediction of admission at the time of emergency health care. The team of Kim [7] used the regular administrative data for creating model using regression technique as well. But the model was weak in accuracy. The results were better when Xie [8] used Coxian Phase model which achieved better performance rate than logistic regression technique. Wang [9] along with the team worked on the model using different type of machine learning algorithms.



The expert opinion of the staff working in ED combined with logistic regression model, naive Bayes algorithm were used by Peck [10] and team mates. After the implementation of the models Peck's team found that logistic model results were different than the expert prediction. The accuracy rate of the technical model was more than the accuracy of experts. Therefore logistic regression technique was generalized [11] and implemented in other hospitals. Implementation of the model by Peck and team [12] showed that the waiting time for the patients was reduced. The ward wise admission forecast was the idea of Qui [13] and group. The team used vector machine for prediction, but the accuracy varied for the different wards.

Machine learning algorithms were used by Lucini [14] and team. The team used eight algorithms out of which six of them showed the similar results. The team of Cameron [15] made an objective test and compared it with the opinion of the staff of ED. The staffs were accurate for the cases where the patients were admitted. There is wide range of methods used in past for the prediction of the patients. The previous work focuses on a small range of data and implementation to only particular cases. There is a requirement of exploring more techniques or combination of techniques in one model. The focus of the developer should be to design a model which can be used at different hospitals. This is possible by use of all the possible basic techniques along with advanced techniques. The main priority for developing and implementation of forecasting prediction models is the best health care and convenience of the patients.

### 3. METHODS

Data mining consist of number of tasks to identify patterns in stored data in emergency departments. The data mining tasks are data extraction data cleansing and feature engineering; data visualisation and descriptive statistics; data splitting into training and test sets; model tuning using the trail mode and 10 fold cross validation repeated 5 times; prediction admissions based on the data set; evaluation of model performance based on the output. The execution of the seven data mining tasks is necessary for separation of extra data from the records to make the prediction more specific and accurate. The final model designed needs to be made with a perspective that it can be implemented in hospitals with different staff numbers, infrastructure and administration with none or few changes.

The data used for analysis consisted of complete information of patients. On the basis of past analysis of methods, data a wide variable range is taken into consideration before the design of final model. The result of analysis can be used in the final model design which consist of variables like hospital location ; date and time of attendance, gender, arrival mode; staff; previous history; time of previous admission ;patient admitted or not. Feature engineering implemented on the attendance resulted in the clarification of time, date, day, week, month of the year. The admission of the patient is the dependent variable in the final module. The analysis of data before design of model helped to exclude the missing data, the direct admissions data and the regular patients who do not follow the path of ED from the records for completion of design for final model.



### 3.1 Machine Learning algorithms and performance

Logistic regression, a decision tree and Gradient boosted machines are the three machine learning algorithms applied for the design of the forecast model. The binary dependent variable is predicted by logistic regression. The examples of binary variables are positive/negative; deceased/alive; or main focus here is on admit/not admit. Use of Logit link function enables the calculations of odd occurring in an outcome. Recursive partitioning technique from RPART combined with decision tree method divides the data in nodes. The outcome contains the most essential variable nodes. Outfitting is excluded by pruning of the outcome tree.

GBM technique is mainly used for boosting the output and enhancing the final decision tree obtained from a group of decision trees. The use of three different algorithms logistic regression as traditional, RPART decision tree and GBM as advanced the comparison of outputs builds an advanced prediction model. The complexity and practical implementation vary for all the three models. Various steps are taken to enhance performance of implementing model and prevent over fitting. The accuracy rate after implementation is different for each method used. Gradient boosting's performance is best when compared to other prediction methods. It should be considered that the final models will be used by the ED staff of various hospitals. For technical students the knowledge of algorithms is easier to understand. In difficulty level GBM is more difficult to understand and implement than the other two methods. Hence the final implementation of prediction model should be understandable by the hospital staff with less difficulty.

### 4. CONCLUSION

The analysis aimed on finding a solution for the difficulties faced by patients and staff at the time of ED crowding. Prediction models designed using data from records in ED differ in accuracy to the patients admitted and the patients predicted. GBM's track record of accuracy is better than logistic regression and decision tree. The implementation and usage of the latter two methods is found to be easier and understandable by non-technical hospital ED staff. There is a variation in every aspect and all the facts are considered important. When the accuracy level of logistic regression combined with decision tree is close to the accuracy of GBM, the hospital might consider the implementation of former's model. The purpose of implementing a prediction model is to provide better health care to the patients. More accurate prediction avoids confusion in staff. The advantage storage of emergency health care medicines and equipment is that hospital staff can effectively attend a large number of patients in situations of crowding. Prediction model's accuracy, personal staff experience, staff attendance, health care protocols all lead to better health care of patients. Hospitals objective are always to provide excellent health care service to all the patients. Hence the necessary step of using advanced technology for prediction model design is implemented.



## REFERENCES

- [1] Byron Graham, Raymond Bond, Michael Quinn and Maurice Mulvenna, Senior Member, IEEE, *Using Data Mining to Predict Hospital Admissions from the Emergency Department*. DOI: 10.1109/ACCESS.2018.2808843.
- [2] M.A. LaMantia, T.F. Platts-Mills, K. Biese, C. Khandelwal, C. Forbach, C.B. Cairns, J. BusbyWhitehead, J.S. Kizer, *Predicting hospital admission and returns to the emergency department for elderly patients*, *Acad. Emerg. Med.* 17 (2010) 252–259. doi:10.1111/j.1553-2712.2009.00675.x
- [3] M.R. Baumann, T.D. Strout, *Triage of Geriatric Patients in the Emergency Department: Validity and Survival with the Emergency Severity Index*, *Ann. Emerg. Med.* 49 (2007) 234–240. doi:10.1016/j.annemergmed.2006.04.011.
- [4]. Boyle, M. Jessup, J. Crilly, D. Green, J. Lind, M. Wallis, P. Miller, G. Fitzgerald, *Predicting emergency department admissions*, *Emerg. Med. J.* 29 (2012) 358–365. doi:10.1136/emj.2010.103531.
- [5] Y. Sun, B.H. Heng, S.Y. Tay, E. Seow, *Predicting hospital admissions at emergency department triage using routine administrative data*, *Acad. Emerg. Med.* 18 (2011) 844–850. doi:10.1111/j.1553-2712.2011.01125.x.
- [6] A. Cameron, K. Rodgers, A. Ireland, R. Jamdar, G.A. McKay, *A simple tool to predict admission at the time of triage.*, *Emerg. Med. J.* 32 (2013) 174–9. doi:10.1136/emermed-2013-203200.
- [7] S.W. Kim, J.Y. Li, P. Hakendorf, D.J.O. Teubner, D.I. Ben-Tovim, C.H. Thompson, *Predicting admission of patients by their presentation to the emergency department*, *EMA - Emerg. Med. Australia's.* 26 (2014) 361–367. doi:10.1111/1742-6723.12252.
- [8] B. Xie, *Development and Validation of Models to Predict Hospital Admission for Emergency Department Patients*, *Life science Glob.* 2 (2013) 55–66
- [9] J. Wang, C.P. Lim, D. Creighton, A. Khorsavi, S.Nahavandi, J. Ugon, P. Vamplew, A. Stranieri, L.Martin, A. Freischmidt, *Patient admission prediction using a pruned fuzzy min max neural network with rule extraction*, *Neural Comput. Appl.* 26 (2015) 277–289. Doi: 10.1007/s00521-014-1631-z.
- [10] J.S. Peck, J.C. Benneyan, D.J. Nightingale, S.A.Gaehde, *Predicting emergency department inpatient admissions to improve same-day patient flow*, *Acad. Emerg. Med.* 19 (2012) 1045–1054. doi:10.1111/j.1553-2712.2012.01435.x
- [11] J.S. Peck, S.A. Gaehde, D.J. Nightingale, D.Y.Gelman, D.S.Huckins, M.F. Lemons, E.W. Dickson, J.C. Benneyan, *Generalizability of a simple approach for predicting hospital admission from an emergency department*, *Acad. Emerg. Med.* 20 (2013) 1156–1163. doi:10.1111/acem.12244
- [12] J.S. Peck, J.C. Benneyan, D.J. Nightingale, S.A. Gaehde, *Characterizing the value of predictive analytics in facilitating hospital patient flow*, *IIE Trans. Healthc. Syst. Eng.* 4 (2014) 135–143. doi:http://dx.doi.org/10.1080/19488300.2014.930765.
- [13] S.Qiu, R.B. Chinnam, A. Murat, B. Batarse, H. Neemuchwala, W. Jordan, *A cost sensitive inpatient bed reservation approach to reduce emergency department boarding times*, *Health Care Manag. Sci.* 18 (2015) 67–85. Doi: 10.1007/s10729-014-9283-1.



- [14] F.R. Lucini, F.S. Fogliatto, G.J.C. da Silveira, J. Neyeloff, M.J. Anzanello, R. de S. Kuchenbecker, B.D. Schaan, *Text mining approach to predict hospital admissions using early medical records from the emergency department*, *Int. J. Med. Inform.* 100 (2017) 1–8. doi:10.1016/j.ijmedinf.2017.01.001
- [15] A. Cameron, A.J. Ireland, G.A. McKay, A. Stark, D.J. Lowe, *Predicting admission at triage: are nurses better than a simple objective score?*, *Emerg. Med. J.* 34 (2017) 2–7. Doi: 10.1136/emered-2014-204455