IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 21, Issue 2, Ser. I (Mar - Apr 2019), PP 15-22
www.iosrjournals.org

Predicting Emergency Department Services in Hospital Using Machine Learning

#1Palakurthi Shashank, #2Annadi Jahnavi, #3Dr.Madugula Sujatha,
B.Tech Student, B.Tech Student, Associate Professor,
Dept Of CSE, Jyothishmathi Institute Of Technology & Science, Karimnagar,T.S., India.
Corresponding Author: palakurthi Shashank

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. To be able to predict, at the time of triage, whether a need for hospital admission exists for emergency department (ED) patients may constitute useful information that could contribute to system wide hospital changes designed to improve ED throughput. The objective of this study was to develop and validate a predictive model to assess whether a patient is likely to require inpatient admission at the time of ED triage, using routine hospital administrative data. Using Single Data Mining Technique hospital admissions for the emergency department has been comprehensively investigated showing acceptable levels of accuracy. The Machine Learning technique for Healthcare in developing algorithms that is used to identify the complex patterns with large amount of data. This technique implies the ways to make intelligent data-driven decisions. 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. The objective of this paper analyzes importance of big data and the various steps involved in machine learning techniques in healthcare.*

Keywords: *Triage Systems, emergency department, hospitals, machine learning, predictive models.*

Date of Submission: 22-02-2019 Date of acceptance: 08-03-2019

I. Introduction

In recent years we have witnessed a dramatic increase of electronic health data, including extensive Electronic Medical Records (EMR) recording patient conditions, diagnostic tests, labs, imaging exams, genomics, treatments, outcomes, claims, financial records, clinical guidelines and best practices etc[1].

Healthcare professionals are now increasingly asking the question: what can we do with this wealth of data? How can we perform meaningful analytics on such data to derive insights to improve quality of care and reduce cost? Healthcare Analytics needs to cover the whole spectrum including both Knowledge Driven Analytics and Data Driven Analytics[2]. Knowledge driven approaches operate on knowledge repositories that include scientific literature, published clinical trial results, medical journals, textbooks, as well as clinical practice guidelines. Traditionally the gold standard of evidence in healthcare has been produced through the randomized controlled trial process.

While most emergency department (ED) visits end in discharge, EDs represent the largest source of hospital admissions. In the ED, patients are first sorted by acuity in order to prioritize individuals requiring urgent medical intervention. This sorting process, called "triage", is typically performed by a member of the nursing staff based on the patient's demographics, chief complaint, and vital signs. Subsequently, the patient is seen by a medical provider who creates the initial care plan and ultimately recommends a disposition, which this study limits to hospital admission or discharge[3,4].

Prediction models in medicine seek to improve patient care and increase logistical efficiency For example, prediction models for sepsis or acute coronary syndrome are designed to alert providers of potentially life-threatening conditions, while models for hospital utilization or patient-flow enable resource optimization on a systems level. Early identification of ED patients who are likely to require admission may enable better optimization of hospital resources through improved understanding of ED patient mixtures. It is increasingly understood that ED crowding is correlated with poorer patient outcomes. Notification of administrators[5] and inpatient teams regarding potential admissions may help alleviate this problem. From the perspective of patient care in the ED setting, a patient's likelihood of admission may serve as a proxy for acuity, which is used in a number of downstream decisions such as bed placement and the need for emergency intervention[6].

Numerous prior studies have sought to predict hospital admission at the time of ED triage. Most models only include information collected at triage such as demographics, vital signs, chief complaint, nursing notes, and early diagnostics, while some models include additional features such as hospital usage statistics and past medical history[7]. A few models built on triage information have been formalized into clinical decision rules such as the Sydney Triage to Admission Risk Tool and the Glasgow Admission Prediction Score[8]. Notably, a progressive modeling approach that uses information available at later time-points, such as lab tests ordered, medications given, and diagnoses entered by the ED provider during the patient's current visit, has been able to achieve high predictive power and indicates the utility of these features. We hypothesized that extracting such features from a patient's previous ED visits would lead to a robust model for predicting admission at the time of triage. Prior models that incorporate past medical history[9] utilize simplified chronic disease categories such as heart disease or diabetes while leaving out rich historical information accessible from the electronic health record (EHR) such as outpatient medications and historical labs and vitals, all of which are routinely reviewed by providers when evaluating a patient. In this work showed that using all elements of the electronic health record can robustly predict in-patient outcomes, a prediction model for admission built on comprehensive elements of patient history may improve on prior models[10].

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.

There are two general categories of algorithms: unsupervised and supervised. Unsupervised machine-learning algorithms are typically used to group large amounts of data. 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. These results are then used to make predictions based on out-of-sample data for which the outcome of interest is not known.

II. Related Work

Health is determined by several factors including genetic inheritance, personal behaviors, access to quality health care, and the general external environment (such as the quality of air, water, and housing conditions). In addition, a growing body of research has documented associations between social and cultural factors and health. The influence of social and cultural variables on health involves dimensions of both time (critical stages in the life course and the effects of cumulative exposure) as well as place (multiple levels of exposure). Big data is created, stored, and disseminated through traditional and mobile Internet, smartphones, smart TV, sensor- and RFID based ubiquitous networks, and social media.

Given a set of clinical cases that act as examples, learning in intelligent systems can be achieved using ML methods that are able to produce a systematic description of those clinical features that unique indicate those conditions. As John Smith, Senior Manager for Intelligent information Systems at IBM Research, says one of the most promising near-term applications is for its use in detecting melanoma. By feeding a computer with many images of the cancer, IBM is planning to teach the system how to recognize features associated with the disease as well as support the physician with text-based medical records concerning diagnosis and treatment protocols. Disease detection and surveillance systems provide epidemiologic intelligence that allows health officials to deploy preventive measures and help clinic and hospital administrators make optimal staffing and stocking decisions.

Studies on factors contributing to LOS have regularly appeared in the literature. One study conducted to determine the factors affecting LOS in public hospitals in Lorestan 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. The t-test, one-way ANOVA, and multifactor regression were used for the analysis. They did not provide any prediction model, because they focused on descriptive analysis based on traditional statistical methods .

Rowan et al. proposed and implemented a software package demonstrating that artificial neural networks (ANNs) could be used as an effective LOS stratification instrument in postoperative cardiac patients.

Blais et al.[11] designed a screening and rating tool to quantify variables related to LOS in a medical psychiatric unit. The findings from this study showed that 25 variables, including patient, illness, and treatment variables, were likely to be related to LOS.

Tu and Guerriere [12] indicated that ANNs can be used as a predictive tool to identify patients at increased risk for prolonged intensive care unit LOS following cardiac surgery. They claimed that the back propagation algorithm had not previously been developed for this area.

Lin et al. [13] explored the prediction of hospital stays for first-time stroke patients in a rehabilitation department by a proportional hazard regression (HR) model. They proposed using the HR model to predict the mean LOS of stroke patients.

Wrenn et al. [14] were able to predict LOS for an emergency department through developing and validating an ANN. The results were promising and showed that ANN can predict a patient's LOS within an average of <1.99 hours. Using a cohort of prospectively identified heart failure patients, Wright et al. [15] found that peripheral edema, chest pain, fatigue, serum albumin, serum sodium at admission and peak creatinine could result in hospital stays longer than six days.

Jiang et al. [16] studied the use of four data mining techniques (logistic regression, neural network, decision tree, and ensemble model) to analyze the inpatient discharge data for average LOS based on input variables. The findings from this research showed that the ensemble model was the best fit, and age and chronic disease were the important predictors. Misclassification and average squared error were used to assess the models. The ensemble model had the lowest average squared error (0.21), and the decision tree had the highest average squared error (0.22).

III. Methodology

EXISTING SYSTEM

In Existing system, the single data mining technique is used to predict hospital admissions from the emergency department. There is no previous research that identifies which data mining technique can provide more reliable accuracy in identifying suitable treatment for hospital admissions from the emergency department. Practical use of hospital database systems and knowledge discovery is difficult in hospital admissions from the emergency department[17].

DISADVANTAGES

1. Hospitals do not provide the same quality of service even though they provide the same type of service.
2. There is no previous research that identifies which data mining technique can provide more reliable accuracy in identifying suitable solution to predict hospital admissions from the emergency department.
3. It takes more time consumption for practical use of hospital database systems.

PROPOSED SYSTEM

In Proposed System, we are applying data mining techniques (Hybrid) in identifying suitable solution to predict hospital admissions from the emergency department. Apply single data mining techniques to predict hospital admissions from the emergency department is benchmark dataset to establish baseline accuracy for each single data mining technique to predict hospital admissions from the emergency department. Apply the same single data mining techniques used in hospital admission to predict hospital admissions from the emergency department dataset to investigate if single data mining techniques can achieve equivalent (or better) results in suitable solution identifying to predict hospital admissions from the emergency department. Apply hybrid data mining techniques to predict hospital admissions from the emergency department benchmark dataset to establish baseline accuracy for each hybrid data mining technique in the to predict hospital admissions from the emergency department. Apply the same hybrid data mining techniques used in hospital admission to predict hospital admissions from the emergency department dataset to investigate if hybrid data mining techniques can achieve equivalent (or better) results in identifying suitable solution identifying to predict hospital admissions from the emergency department[18].

ADVANTAGES

1. By applying data mining techniques to help emergency department in hospital to predict hospital admissions from the emergency department.
2. Hybrid data mining techniques are used for selecting the suitable to predict hospital admissions from the emergency department.
3. Time consumption is less.
4. High Performance and Accuracy.

In this segment different existing methods has been talked about. Distributed storage is viewed as an arrangement of dispersed server farms that for the most part Utilize virtualization innovation and supplies interface for information stockpiling.

IV. Data Mining Algorithms

Finding undiscovered information and useful patterns in a database is often referred to as data mining. Data mining is heavily used in the health and medical field in applications such as disease prediction and patient management[18]. Relationships, rules, and essential information about or from the data cannot be easily extracted because of database size and other features. We used some of the most common predictive data mining methods for our goals as follows.

1) Artificial neural networks

ANNs are used to perform multivariate analysis to identify both linear and non-linear patterns among data variables. Due to their good predictive performance, ANNs are the most popular method in various areas of medicine and lead to appropriate decisions. An ANN consists of many connected processing elements, including multiple input nodes and weighted interconnections. The radial-basis-function (RBF) ANN was developed to recognize CAD[20].

2) Support vector machines (SVMs)

A category of classification that has received increasing attention in recent years is the SVM[21]. It is a new method for classification of both linear and non-linear data, and in terms of predictive accuracy, it is a powerful algorithm. In fact, SVM is a linear learning machine constructed through an algorithm that uses an optimization criterion. We apply RBF kernel mode because of its good general performance and because it has the smallest number of parameters.

3) Decision tree

Generally, a decision tree as a visual and analytical decision support tool is a graphic representation of obtained knowledge in the form of a tree (flow chart like structure), where each non-leaf node denotes a test on an attribute, and each branch indicates an output of the test. It uses a combination of mathematical and computational techniques to aid description and classification, and to extract knowledge of data set. Because nodes and branches are organized hierarchically, they are easy to understand and interpret[22]. They are reliable and have better accuracy in clinical decision-making. C5.0 decision trees are the most current decision tree algorithms. The C5.0 algorithm with 10-fold cross validation and 20 trials using boosting was applied in this research.

4) Ensemble models

The ensemble method creates a new model by combining SVM, C5.0, and ANN models.

V. Machine Learning Techniques In Healthcare

Machine learning (ML) provides methods, techniques and tools that can help solving diagnostic and prognostics problems in a variety of medical domains. ML is being used for the analysis of the importance of clinical parameters and of their combinations for prognosis, eg. prediction of disease progression, for the extraction of medical knowledge for outcomes research, for therapy planning and support, and for overall patient management.

5.1 Types of Machine Learning Algorithms

Four different types of machine learning algorithms are available that can be organized into taxonomy based on the desired outcome of the algorithm or the type of input available for training the machine. Thompson noted, "The terminology used in machine learning is different than that used for statistics. For example, in machine learning, a target is called a label, while in statistics it's called a dependent variable." [23] The key types of machine learning include:

Supervised learning.
Unsupervised learning.
Semi supervised learning.
Reinforcement learning.

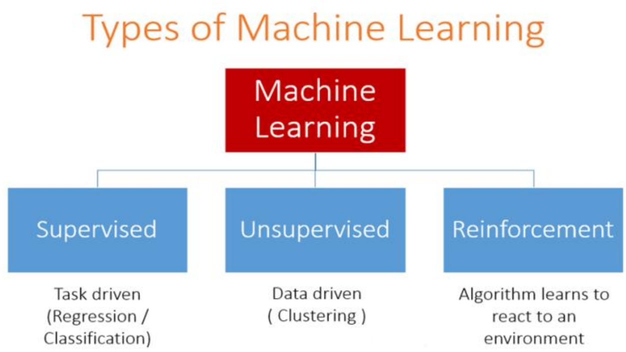

Figure 1. Types of Machine Learning

Supervised learning is a type of machine learning that uses a known dataset (called the training dataset) to make predictions. The training dataset includes input data and labelled response values. Supervised machine learning techniques are more suitable for medical data classification.
Unsupervised learning is a type of machine learning used to draw inferences from datasets consisting of input data without labelled responses.

5.2 Applications of Machine Learning Techniques in Health Care

Machine learning algorithms are effective in recognizing complex patterns within rich and massive data. This capability is particularly well-suited to medical applications, especially those that depend on complex proteomic and genomic measurements. As a result, machine learning is frequently used in various disease diagnosis and detection. In clinical applications machine learning algorithms can produce better decisions about treatment plans for patients by means of providing effective healthcare system

5.2.1 Discrete Event Simulation

Health care organizations are using this technique to predict wait times for patients in emergency department waiting rooms. The models use factors such as staffing levels, patient data, emergency department charts, and even the layout of the emergency room itself to predict wait times **5.2.2 Free-text physician notes**

IBM researchers have found a way to extract heart failure diagnosis criteria from free-text physician notes method. They developed a machine learning algorithm that combs through physicians free-form text notes (in the electronic health records) and synthesize the text using a technique called “Natural Language Processing” (NLP). Similar to the way a cardiologist can read through another physician’s notes and figure out whether a patient has heart failure, computers can now do the same

5.2.3 Predicting Strokes and Seizures-Singapore based start-up Healint launched an app called JustShakeIt that enables a user to send an emergency alert to emergency contacts and/or caregivers simply by shaking the phone with one hand. The program uses a machine learning algorithm to distinguish between actual emergency shakes and everyday jostling. In addition to the JustShakeIt app, Healint is working on a model that analyzes patients’ cell phone accelerometer data to help identify warning signs for chronic neurological conditions.[24]

5.2.4 Proprietary predictive model

Using this predictive model, hospitals can predict emergency room admissions. Thus the application of machine learning may benefit patients either by reducing costs, improving accuracy, or disseminating expertise that is in short supply.

5.2.5 Machine Learning Techniques in Numerous Disease Predictions and Diagnosis:

Machine learning plays a key role in many radiology applications. Machine learning identifies complex patterns automatically and helps radiologists make intelligent decisions on radiology data such as conventional radiographs, CT, MRI, and PET images and radiology reports [25].

Steps to apply Machine Learning in Health Care Data

1. Define the Problem-Describe the problem informally and formally and list assumptions and similar problems. List the technique for solving the problem, the benefits a solution provides and how the solution will be used with health care system.

2. Select Data and Prepare a model-Data preparation with a data analysis phase that involves summarizing the attributes and visualizing them using scatter plots and histograms.

Step 1: Data Selection: Consider what data is available, what data is missing and what data can be removed.
Step 2: Data Pre-processing: Organize your selected data by formatting, cleaning and sampling from it.
Step 3: Data Transformation: Transform preprocessed data ready for machine learning by engineering features using scaling, attribute decomposition and attribute aggregation.

3. Spot Check Algorithms – Test and validate the process. Spot checking algorithms is a part of the process of applied machine learning. On a new problem, we need to quickly determine which type or class of algorithms is good at picking out the structure in the problem and which are not. Loading up a bunch of standard machine learning algorithms into the data set harness and performing a formal experiment. There are three key benefits of spot-checking algorithms in machine learning problems are speed, objective and result..

4. Improve Result- If better performance is needed, it becomes necessary to utilize more advanced strategies to augment the performance of the model .Using Machine Learning techniques it will reduce the variance of the performance measure. The process of improving results involves:

- ✓ Algorithm Tuning: where discovering the best models are treated like a search problem through model parameter space.
- ✓ Ensemble Methods: where the predictions made by multiple models are combined.
- ✓ Extreme Feature Engineering: where the attribute decomposition and aggregation seen in data preparation is pushed to the limits.

5. Apply Result -Depending on the type of problem to solve, the presentation of results will be very different. There are two main facets to making use of the results of your machine learning endeavor:

- ✓ Report the results
- ✓ Operationalize the system

After these steps have been completed, if the model appears to be performing satisfactorily, it can be deployed for its intended task. The model may be utilized to provide score data for predict the disease, for projections of Electronic Medical Record, to generate useful insight for decision making or research, or to automate tasks. Machine learning is closely related to computational statistics, a procedure which focuses in prediction through the use of computers. ML methods are implemented with optimization techniques, which deliver methods, theory and application domains to the field.

VI. Expected Results

Emergency Department and surveillance systems provide epidemiologic intelligence that allows health officials to deploy preventive measures and help clinic and hospital administrators make optimal staffing and stocking decisions, The below figures show the results of hospital emergency department.

Fig.2. Graph for Total Admitted Patient Details

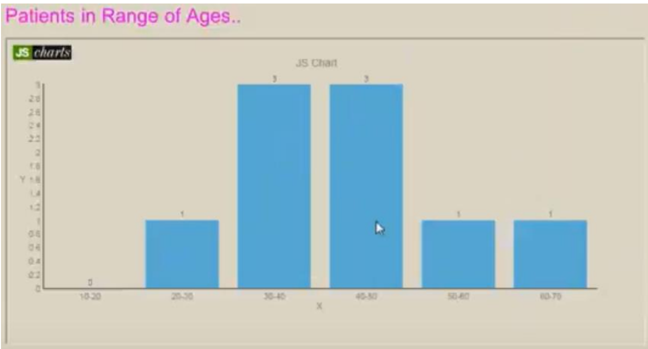

Fig.3. Graph for Admitted Patient Age Details

VII. Conclusion

With more and more data available, machine learning techniques are becoming increasingly popular as they get better at looking at massive amounts of data. The most important challenges in clinical practice and biomedical research include the need to develop and apply novel tools for the effective integration, analysis and interpretation of complex biomedical data with the aim to identify testable hypothesis, and build accurate models. Big data Analytics gives a great boost to leverage the benefits of chaotic environment in healthcare. Using these new techniques it is easier to develop therapics and products. By comparing the effectiveness of machine learning models are used in health care delivery and services and assessing health. Proposed model is a perfect match for big data with healthcare since cloud computing provides unlimited resource on demand. This allows aggregation of multiple disparate workloads with varying performance goals into very large clusters. Incorporating big data in healthcare clearly has the ability to transform the industries. Personalized medicines and early monitoring and diagnosis are also expected due to the appropriate analysis of big data.

References

[1]. Wright SP, Verouhis D, Gamble G, Swedberg K, Sharpe N, Doughty RN. Factors influencing the length of hospital stay of patients with heart failure. Eur J Heart Fail. 2003; 5(2):201–209.

[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]. Robinson GH, Davis LE, Leifer RP. Prediction of hospital length of stay. Health Serv Res. 1966; 1(3):287–300.

[10]. 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.

[11]. Blais MA, Matthews J, Lipkis-Orlando R, Lechner E, Jacobo M, Lincoln R, et al. Predicting length of stay on an acute care medical psychiatric inpatient service. Adm Policy Ment Health. 2003; 31(1):15–29.

[12]. Tu JV, Guerriere MR. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. Proc Annu Symp Comput Appl Med Care. 1992; 31(1):666–672.

[13]. Lin CL, Lin PH, Chou LW, Lan SJ, Meng NH, Lo SF, et al. Model-based prediction of length of stay for rehabilitating stroke patients. J Formos Med Assoc. 2009; 108(8):653–662.

[14]. Wrenn J, Jones I, Lanaghan K, Congdon CB, Aronsky D. Estimating patient's length of stay in the Emergency Department with an artificial neural network. AMIA Annu Symp Proc. 2005; 2005:1155.

[15]. Stoskopf C, Horn SD. Predicting length of stay for patients with psychoses. Health Serv Res. 1992; 26(6):743–766.

[16]. Negassa A, Monrad ES. Prediction of length of stay following elective percutaneous coronary intervention. ISRN Surg. 2011; 2011:714935.

[17]. Jilani TA, Yasin H, Yasin M, Ardil C. Acute coronary syndrome prediction using data mining techniques: an application. Int J Inf Math Sci. 2009; 5(4):295–299.

[18]. Liu P, Lei L, Yin J, Zhang W, Naijun W, El-Darzi E. Healthcare data mining: predicting inpatient length of stay. In : Proceedings of the 3rd International IEEE Conference Intelligent Systems; 2006 Sep 4-6; London, UK. p. 832–837.

[19]. Kudyba S, Gregorio T. Identifying factors that impact patient length of stay metrics for healthcare providers with advanced analytics. Health Informatics J. 2010; 16(4):235–245.

[20]. Rani KU. Analysis of heart diseases dataset using neural network approach. Int J Data Min Knowl Manag Process. 2011;

[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.

[24]. Suryawanshi RD, Thakore DM. Classification techniques of datamining to identify class of the text with fuzzy logic. In : Proceedings of 2012 International Conference on Information and Computer Applications; 2012 Feb 17-18; Hong Kong. p. 263–267.

[25]. Sitar-Taut DA, Sitar-Taut AV. Overview on how data mining tools may support cardiovascular disease prediction. J Appl Comput Sci. 2010; 4(8):57–62.

AUTHOR'S PROFILE:

[1].PALAKURTHI SHASHANK Pursuing his B.Tech Final Year in Computer Science and Engineering Department at Jyothishmathi Institute Of Technology & Science, Karimnagar ,T.S., India.

[2]. ANNADI JAHNAVI Pursuing her B.Tech Final Year in Computer Science and Engineering Department at Jyothishmathi Institute Of Technology & Science, Karimnagar ,T.S., India.

[3]. Dr. MADUGULA SUJATHA Working as Associate Professor in Department of Computer Science and Engineering at Jyothishmathi Institute Of Technology & Science, Karimnagar ,T.S., India. Her Interested areas are Data Mining, Machine Learning and Artificial Intelligence.

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

palakurthi Shashank. " Predicting Emergency Department Services in Hospital Using Machine Learning " IOSR Journal of Computer Engineering (IOSR-JCE) 21.2 (2019): 15-22.