



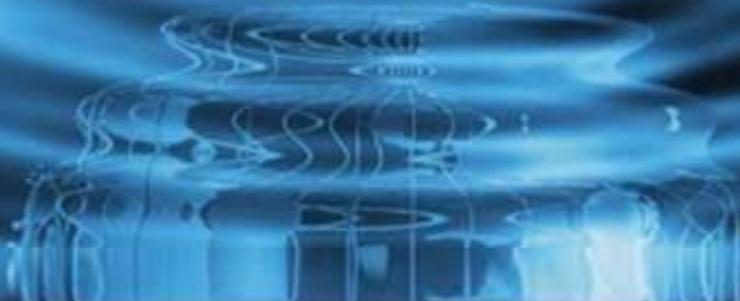
**IJITCE**

**ISSN 2347- 3657**

# **International Journal of**

## **Information Technology & Computer Engineering**

[www.ijitce.com](http://www.ijitce.com)



**Email : [ijitce.editor@gmail.com](mailto:ijitce.editor@gmail.com) or [editor@ijitce.com](mailto:editor@ijitce.com)**

## **A PREDICTION MODEL FOR THE PATIENTS ADMISSIONS WITH MACHINE LEARNING**

**N.Jagannadha Rao, Y. Ramesh Kumar, S. Lakshmi Aparna, Pothumudi Manasa, Srikanth Kosaraju**

**Abstract:** people will face many problems in Hospitals while taking Admission. If it is in a popular hospital, they should wait hours together to take just admission. But it is not at all good at Emergency Department. Very serious cases will admit in Emergency Department. So, we need to use more innovation technique to ameliorate patient flow and prevent Overflowing. So, data mining techniques will show us a pleasant method to predict the ED Admissions. Here we Analyzed an algorithm for predicting models i.e., Naive Bayes, Random Forests, Support Vector Machine. For the prediction we should identify a handful of factors associated to Hospital admission including age, gender, systolic pressure, diastolic pressure, diabetes, previous records in the preceding month or year, admission. We also say about the algorithms which we used in detail. We use Random Forests algorithm for classifying the data into categories for improving the accuracy of prediction. Naive Bayes is used to identify the probabilities for each attribute and helps in predicting the outcome. Support Vector machine is used to classify the given input particular category which helps in predicting the outcome.

### **1. INTRODUCTION**

One of the biggest yet overlooked problems in the Medical Industry is Emergency Department Crowding. These are the most severely injured or patients who need immediate attention. However, it is often very difficult to identify the state of all the patients in the Emergency ward which leads to making wrong decisions which soon leads to overcrowding. This is why the ability to identify the state of a patient has become crucial worldwide.

Overcrowding might seem like an easy problem to get over but in reality, it is very hard to handle. The consequences are harsh and will directly impact the patients as well as the staff in the hospital as the wait times will increase drastically and it will be too late for anyone to react due to the shortage of required staff. This is why it is necessary for us to come up with innovative approaches to solve this global issue to improve the patient flow and prevent patient crowding.

---

Department of CSE, PRAGATI Engineering College (Autonomous), Surampalem, A.P, India.

---

One of the best approaches to this method over the past few years has been the use of datamining using various ML techniques in order to predict the state of various emergency patients that are currently admitted in the hospital. However, there are a few cases in which emergency crowding takes place due to the shortage of doctors or even the lack of inpatient beds. These are mainly caused due to the fact that the patients from the emergency ward are

transferred to these inpatient beds. This is one of the problems we can easily rectify with the help of data mining in order to identify patients that are inpatient admissions from those who are not so that we can avoid any confusions in our system. In this study we will mainly focus upon implementing various machine

learning algorithms and developing models in order to predict the state of the patients that are being admitted into the emergency department. We will also be comparing the performance of our model with a few various approaches that are already in the world. Patients who plan on visiting the hospitals for various issues and those that are in the emergency department will be required to go through various phases between the time that they arrive to their time of discharge. In these phases will focus upon the various decisions that they had to make depending upon their previous phases. During these phases we will collect various data from the patients such as their patient's age, gender, systolic pressure, diastolic pressure, diabetes, previous records based on these factors the patient will be admitted.



Fig:1 Age groups

The emergency attendees may come through main reception or in ambulance at this point of time depending on the situation of the patient the details should be taken for some of

the medications age, gender, blood pressure, diabetes plays a vital role for the further treatment. Usually, to collect the data from the patient it takes ten to fifteen minutes the patient who comes with emergency may not have time to complete all this procedure. To identify such cases, we must use a Triage Scale in order to understand the condition of the patient and how urgently they require medical care. This is one of the most important phases for the safety of any patient. When we look into the previous records of any hospital, we can clearly identify that there were far more aged people admitted in the hospital when compared to children or adults. This has caused chaos at emergency departments due to a lack of knowledge. Blockchain is the main tool to facilitate this need and when combined with different hashing techniques, this becomes a powerful method for protecting the data. It also helps in eliminating the need for constant verification of certificates. Blockchain technology is used to reduce the incidence of certificate forgeries and ensure that the security, validity, and confidentiality of graduation certificates would be improved. Technologies that exist in security domains include digital signatures, which are used in digital documents to provide authentication, integrity, and non-repudiation. Also, with block chain in play, the storage of certificates

is more secure. With these technologies, an application created that facilitates the secure validation of digital certificates.

regarding the procedures and department medical systems. The number of visit rates to a hospital has been rising rapidly over the past decade. Due to this it is essential for us to create a quick and accurate Triage System in order to assess all the patients. Once the patient has undergone the Triage Process they will shift to the clinical room where they will be consulted by a clinician who will provide the best course of action for the patient. There are various Triage Systems that are used

commonly around

the world. However, the two most commonly used triage systems are those that use either a 3 Level Triage System or a 5 Level Triage System. 3 Level

Triage System

labels patients as Emergent, Urgent and Non urgent from the highest to lowest level respectively. Similarly, the 5 Level System is broken down as Resuscitation, Emergent, Urgent, Less Urgent, Not Urgent from lowest level to highest level respectively.

Various studies around the world have shown that the 5 Level Triage System has been more reliable than the standard 3 Level System. It has done a better job in predicting the consumption of resources, length of stay, admission rates and mortality. Building a Triage System that is highly accurate and precise can play a major impact in the medical industry as it could save millions of lives. Our

Study is based upon two major objectives. Our first objective is to create and develop a model that is able to accurately predict whether a patient from the emergency department will be admitted into the hospital. Our later objective is to study the performance of various other machine learning algorithms in this sector. In order to predict the state of a patient we must first have our heads wrapped around the

knowledge of various mathematical models. The previous research was done by using logistic regression, decision tree and time series forecasting algorithms. In the previous analysis when compared with other algorithms like logistic regression, decision tree and gradient boosted. Gradient boosted got them more accurate as we use decision tree it is not suitable for longer data sets and need to perform pruning in decision trees whereas in gradient boosted it merges the weaker trees and forms the stronger one which helps in the prediction. According to the statistics the rate of patient stays, or visits was gradually increased from the year 2005 to 2014. Annual average growth rate for inpatient stay was 5.7% and cumulative increase was 64.1% whereas in Emergency Department visits annual average growth rate was 8.0% and cumulative increase was 99.4%. Objective is to find the model which suits the best and gives the accurate results for predicting the admission in the emergency department. Here the comparison of three machine learning algorithms was done (i.e.) Naïve Bayes, Support vector machine (SVM), Random Forest classifier. After comparison Support Vector machine got the most accurate results when compared to others.

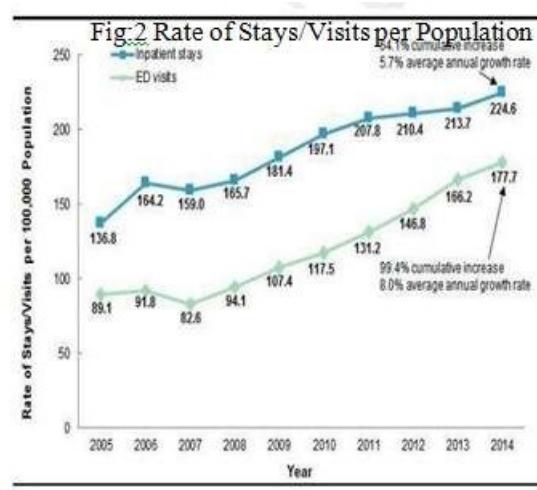


Fig.2 Rate of Stays/Visits per Population

## 2. LITERATURE REPORT

## **Content Extraction Studies using Neural Network and Attribute Generation**

The amount of information available on web today is more than at any point in history, and greater challenges arouse due to this huge wealth of information available. Also, to deal with this information overload, challenging tools are required. Method of Analysis: Internet in the present day especially in India is spreading both in rural and urban areas. Bilingual and Multilingual websites are increasing to a larger extent. Even websites are becoming multitasking. Our main problem is to deal with multilingual web documents and ancient documents. Because, content extraction becomes difficult when such documents are considered. The present paper proposes a neural network approach and attribute generation to justify the content extraction studies for multilingual web documents. Findings: Results obtained are well defined and a thorough analysis is done. Novelty/Improvement: The method is versatile in using pixel-maps, analytically stable in that the matrix input is used and is demonstrated for adoption to different models.

### **Impact of streaming "fast track" emergency department patients**

Fast track systems to stream emergency department (ED) patients with low acuity conditions have been introduced widely, resulting in reduced waiting times and lengths of stay for these patients. We aimed to prospectively assess the impact on patient flows of a fast-track system implemented in the emergency department of an Australian tertiary adult teaching hospital which deals with relatively few low acuity patients.

Methods: During the 12-week trial period, patients in Australasian Triage Scale (ATS) categories 3, 4 and 5 who were likely to be discharged were identified at triage and assessed and treated in a separate fast track area by ED medical and nursing staff rostered to work exclusively in the area.

### **A Comparative Study on Feature Selection in Text Categorization**

This paper is a comparative study of feature selection methods in statistical learning of text categorization. The focus is on aggressive dimensionality reduction. Five methods were evaluated, including term selection based on document frequency (DF), information gain (IG), mutual information (MI), a  $\chi^2$  -test (CHI), and term strength (TS). We found IG and CHI most effective in our experiments. Using IG thresholding with a nearest neighbor classifier on the Reuters corpus, removal of up to 98% removal of unique terms actually yielded an improved classification accuracy (measured by

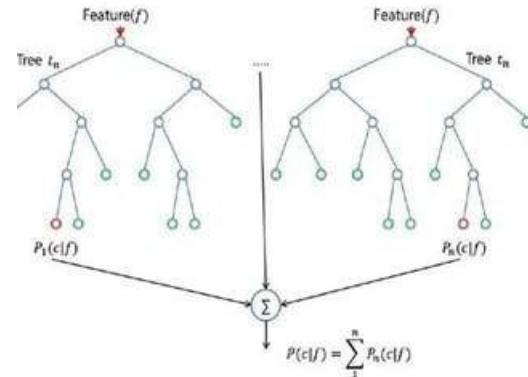
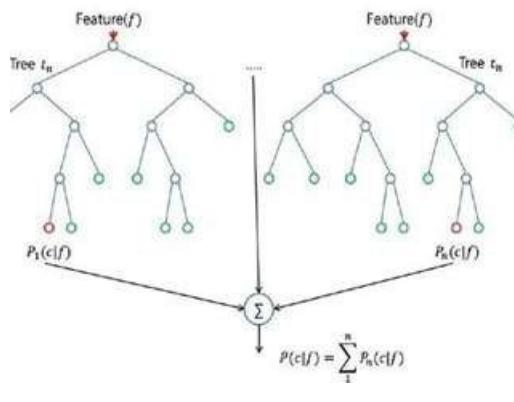
average precision). DF thresholding performed similarly. Indeed we found strong correlations between the DF, IG and CHI values of a term. This suggests that DF thresholding, the simplest method with the lowest cost in computation, can be reliably used instead of IG or CHI when

the computation of these measures are too expensive. TS compares favorably with the other methods with up to 50% vocabulary reduction.

### **department patient flows:**

#### **application of Lean Thinking to health care**

To describe in some detail the methods used and outcome of an application of concepts from Lean Thinking in establishing streams for patient flows in a teaching general hospital ED. Methods: Detailed understanding was gained through process mapping with staff followed by the identification of value streams (those patients likely to be discharged from the ED, those who were likely to be admitted) and the implementation of a process of seeing those patients that minimized complex queuing in the ED.



Results: Streaming had a significant impact on waiting times and total durations of stay in the ED. There was a general flattening of the waiting time across all groups. A slight increase in wait for Triage categories 2 and 3 patients was offset by reductions in wait for Triage category 4 patients. All groups of patients spent significantly less overall time in the department and the average number of patients in the ED at any time decreased. There was a significant reduction in number of patients who do not wait and a slight decrease in access block.

### 3. THEORETICAL ANALYSES

#### Random Forest

Random forest is a commonly used tool in the construction of Decision trees. Instead of following the normal routine it takes a subset of variables and observations in order to construct the decision tree. It builds various decision trees and merges them together in order to form a single decision tree that has high accuracy and prediction. The Random Forest is generally viewed upon as a black box as its predictions are highly accurate. Most people don't bother about the background calculations due to its high accuracy rate. Although we won't be able to change the methods of calculations for the Random Forest, it has a few modifiable factors which can in turn effect the performance of the model or the resources and time balance. We will talk about their variable factors further on in the construction of our Rainforest Model.

Fig:3 Random Forest

#### Parameters to Tune Random Forests

There are 2 major roles that the parameters play when we construct our decision tree. The parameters can either effect the prediction power of our model or they allow us to train our model far more simply. Let's take a look into these parameters in a far more detailed manner. Major concept behind the

Rectified further down the line there shouldn't be any problem with using these libraries on your data.

d

#### **PSEUDO CODE FOR RANDOM FOREST**

F 1. Assume that  $N$  is the no of cases in our training set. Once you've done these take a random sample amongst these  $N$  cases at random and with replacement.

o 2. Take the number of input features or  $t$  variables as  $M$ . We must then specify a number  $n$  such that  $n < M$  and also that there are  $m$  input variables selected at each node.

o  $M$  is then further used in order to split the nodes and known as bestsplit. As we further construct our forest the value  $m$  will remain

1 constant

3. If pruning does not occur each tree is allowed to grow as large as possible.

4. All the constructed trees are merged together to form a single tree to create predictions with much higher accuracy. Majority Voting is the

#### **Advantages of Random Forest**

1. Random Forest algorithm deals with both classification and regression tasks.

2. The Random Forest Model is able to handle any missing values in our data set and still maintain predictions with high accuracy.

3. If there are more trees in the model, the algorithm would not overfit the model.

#### **Disadvantages of Random Forest**

1. Good at classification concept but not so good at regression.

2. As the inner calculations of the Random Forest Model are scarcely known we have very little control on how the model functions.

#### **Applications**

1. They are used within banking sectors in order to segregate loyal and fraud clients.

2. It is used within the medical industry in order to identify the possible combinations of various components in order to validate the correct medicine.

3. It is further used in order to label patients with their respective problems by looking

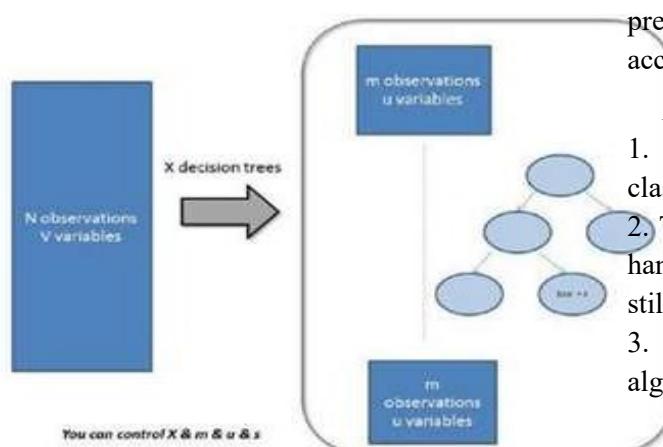


Fig:4

FIG:4 Features which make predictions to the model better the chance of bumping into impurities are relatively high when it comes to using random rainforest libraries. However, they all come with their drawbacks when you keep data interpretation in mind. A simple example for this is using correlation analysis. A feature that typically has a very strong score can be depicted as a low score feature within rainforest. Another recurring impurity is how certain methods are biased toward specific features. However as long as you are able to keep these drawbacks in mind and have them

into their previous records.

4. It can be used in order to identify the behavior of the stockmarket.
5. It is used in image and voice classification.

### Support Vector Machine

The Classification of Linear as well as Non-Linear Data can be simply completed with the help of a Support Vector Machine(SVM). Let 's take a simple look at how SVM 's function or work.SVM 's applies nonlinear mapping in order to convert the original training data into training data in higher dimensions. Once we have established this new dimension the model will begin searching for linear optimal separating hyperplanes. The SVM is able to find and separate these hyperplanes with the usage of support vectors and margins. We will look deeper into these concepts later on in our study. However, in the past decade SVM's have

been attracting a lot of attention. SVM's were first introduced into the picture when Vladimir Vapnik along with his colleagues Bernhard Boser & Isabelle Guy decided to write a paper on them in 1992. Although these group of researchers were the first to have written a paper on SVM's the concept has dated back to the 1960s. SVM's follow a rather complicated internal structure and the time to train them is extremely slow. However, putting this aside, you will be

able to expect outputs which are highly accurate and precise. Another key factor to using SVM's is their ability to be prone to overfitting. A commonly used application of SVM's has been numeric or alphanumeric prediction as well as classification. Other applications for SVM's have included areas such as handwritten language or digit detection, speaker identification, object detection, and Benchmark time series. SVM's are mostly based upon the concepts of decision planes that have predefined boundaries. A decision plane can simply be defined as a barrier that separates the various objects that belong to different membership classes. Let's try to take a look at this simple schematic

example in which objects either belong to the left class or the right class. The line in the middle acts as the boundary or you can say decision plane which separates the right and left class. All the objects that are situated to the left of this line are known as the left class while all those to the right are classified as the right class. When a new object enters into the scenario it falls upon the boundary line which will then make the classification to either push it left or right into its respective class

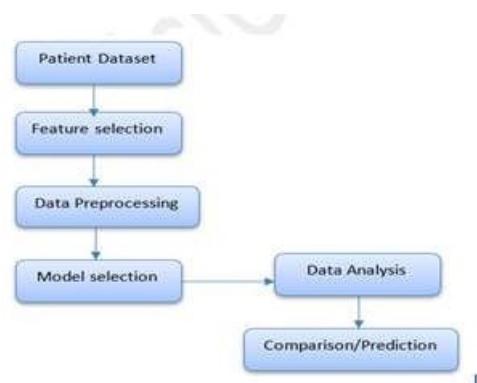


Fig: 5

The schematic example that we have looked into above is nothing more than a basic linear classifier. A linear classifier is nothing more than a classifier that separates the objects into various groups. In our example these groups were color based of Red & Green objects. This is a very basic and simple method of classification. However not all classifications are as simple as this one, they are often far more complex and require more classifications in order to properly segregate the training objects. Let us take a look at another classification with the same segregation of Red & Green objects below.

Compared to our previous linear classification we can clearly see that

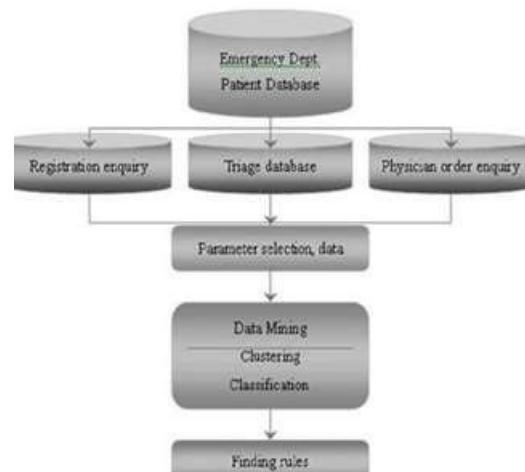
the separation of the objects now requires a curve then a line. A curve is a far more complex structure than a line.

The classification for this curve takes place by drawing various separating lines in order to identify and distinguish the objects from one another. This type of classification is commonly known as a hyperplane classifier and SVM's are the best models in order to handle these types of classifications and tasks.

#### 4. DATA MINING

When we sort through large data sets in order to identify various patterns and establish relationships it is known as Data Mining. These patterns and Relationships can be further used in order to solve various problems through data analytics. Enterprises are able to make predictions upon future trends with the help of Data Mining tools. We are able to do so by using massive amounts of data in order to identify the various patterns and trends. It typically consists of Data Transformation, Pattern Evaluation, Data Cleaning, Pattern Discovery, Data Integration, and Knowledge Presentation. We use Association rules within data mining by exploring and analyzing the data for various if/then patterns. From here we will use various support and confidence criteria in order to form various important relationships among data. Support is defined as the number of times a specific query is found within a database, while confidence is the probability that the if/then case is accurate. There are other parameters used within data mining such as Sequence or Path Analysis, Clustering and Forecasting, Classification, and Sequence or Path Analysis. An ordered list of a set of items is known as a Sequence. It is commonly found in any sort of Database. The Classification Parameter is used in order to detect new patterns. It may also change the structure of our organized data. All Data Mining

techniques are executed in a specific organized manner or flow. For you to get a better understanding have a look at the flowchart below.



#### PROPOSED SYSTEM

In proposed application to predict the ED Admissions. Here we Analysed an algorithm

for predicting models i.e., Naive Bayes, Random Forests, Support Vector Machine. For the prediction we should identify a handful of factors associated to Hospital admission including age, gender, systolic pressure, diastolic pressure, diabetes, previous records in the preceding month or year, admission. We also say about the algorithms which we used in detail. We use Random Forests algorithm for classifying the data into categories for improving the accuracy of prediction. Naive Bayes is used to identify the probabilities for each attribute and helps in predicting the outcome. Support Vector machine is used to classify the given input particular category which helps in predicting the outcome.

#### 5. OUTPUTS

In above screen home page for Patient Admission Prediction, In the above Screen Patient Data set was loaded.



In above screen our application is predicting the patient's admission.



In above screen it displaying about Model comparison

## 6.CONCLUSION

Our study focused upon the advancement and the correlation of various machine learning models that are used in order to look over hospital admissions dealing with the Emergency department. Each model that we looked into was generated using information gathered from various emergency departments. These 3 models were able to be constructed using 3 different techniques which were namely Naïve Bayes, Random Forest Classifier, and Support Vector Machine. Out of the 3 models that we were able to analyze we found that the model which was generated using the SVM classifier was found to be more successful and accurate when compared to the other two models which were generated using Random Forest and Naïve Bayes. The 3 models that we had decided to look into all showed very similar and comparable results. We believe that these models can help many hospitals in facing the global problem of the overflow of patients in the Emergency Departments. They can also help us to increase the Patient Flow in hospitals and reduce crowding overall. We also believe that such models can be used in various other fields in the real world as well in order to monitor the performance of various objects. There is so much we can use these models for in the real world and we believe that we can build upon these models for various use cases.

## 7.REFERENCES

- [1] Li JYZ, Yong TY, Bennett D Et al. Outcome of the interposition of an acute assessment unit in the general medical service of a tertiary teaching hospital. *Med. J. Aust.* 2010; 192:384–7.
- [2] Prakash, K.B. & DoraiRangaswamy, M.A. 2016, "Content extraction studies using neural network and attribute generation", *Indian Journal of Science and Technology*, vol.9,no.22,pp.1-10.
- [3] O'Brien D, Williams A, Blondell Ket al. Impact of streaming 'fast track' emergency department patients. *Aust.Health Rev.* 2006; 30: 525–32.
- [4] King DL, Ben-Tovim DI, Bassham J. Redesigning emergency department patient flows: application of Lean Thinking to health care. *Emerg.Med.Australas.* 2006; 18: 391–
- [5] Gardner RL, Sarkar U, Maselli JH et al. Factors associated with longer ED lengths of stay. *Am. J. Emerg.Med.* 2007; 25: 643–50.
- [6] Prakash, K.B. 2018, "Information extraction in current Indian web documents", *International Journal of Engineering and Technology(UAE)*, vol. 7, no. 2, pp. 68-71.
- [7] Emergency Department Overcrowding in Massachusetts. Making Room in our Hospitals. Issue Brief. The Massachusetts Health Policy Forum, No 12; 2001.
- [8] National Hospital Ambulatory Medical Care Survey. 2002 Emergency Department Summary. Advance Data Number 340. 35pp. (PHS) 2004-1250.
- [9] Prakash, K.B., Kumar, K.S. & Rao, S.U.M. 2017, "Content extraction issues in online web education", Proceedings of the 2016 2nd

International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT 2016, pp. 680.

[10] K.S.S. Joseph Sastry& T. Gunashekhar,‘A Systematic Access Through Machine Learning Methods For Expectation In Malady Related Qualities‘, International Journal of Engineering and Advanced Technology (IJEAT), Volume- 8, Issue6S, August 2019, ISSN: 2249 – 8958

[11] Anusha M, K Karthik, P Padmini Rani &VSrikanth,‘ Prediction of Student Performance using Machine Learning‘, International Journal of Engineering and Advanced Technology (IJEAT), Volume- 8, Issue-6, August 2019, ISSN: 2249 – 8958

[12] Prakash, K.B., Kumar, K.S. &Rao, S.U.M.2017,"Content extraction issues in online web education", Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT 2016, pp. 680.

[13] K. SripathRoy,K.Roopkanth, V.UdayTeja, V.Bhavana, &J.Priyanka,‘ Student Career Prediction Using Advanced Machine Learning Techniques‘,International Journal of Engineering & Technology, 7 (2.20) (2018) 26-29

[14] KVSN Rama Rao, Sivakannan S, M.A.Prasad, R.AgileshSaravanan, \_Technical challenges and perspectives in batch and stream big data machine learning‘, International Journal of Engineering & Technology, 7 (1.3)(2018) 48-51

[15] Reddy, L.S.N.a, Kiran, K.S.a, Brahmani, K.N.a, Vamsidhar, E.a,b,‘ Performance analysis on human activity detection using KNN and random forest(Article)‘, International Journal of Innovative Technology and Exploring Engineering,Volume 8, Issue 7, May 2019, Pages 2817-2821

[16] Prakash, K.B. 2017, "Content extraction studies using total distance algorithm", Proceedings of the 2016 2nd International Conference on Applied and Theoretical

Computing and Communication Technology, iCATccT 2016, pp. 673.