Prediction Of Hospital Admission Using Machine Learning

P.Ganesh(2095A1221)
Informaion Technology
Institute of aeronautical engineering
Hyderabad,India

A.Nithin Reddy(20951A1253)
Information Technology
Institute of aeronautical engineering
Hyderabad,India

S.Rohith(20951A1279)
Information Technology
Institute of aeronautical engineering
Hyderabad,India

Ms G Lohitha
Information Technology
Institute of aeronautical engineering
Hyderabad,India

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 n eed 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 preventing patient crowding. One of the best approaches to this method over the past few years has been the use of data mining 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.
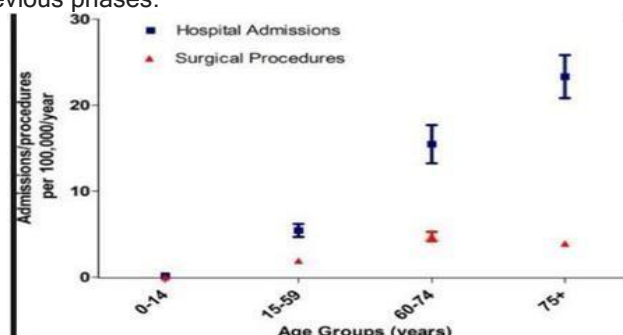

Fig:i

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 shifted to the clinical room where the 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. A 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 showed that the 5 Level Triage System has been far 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 weather 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 the 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 impatient stay was 5.7% and cumulative increase was 64.1% where as 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

2.THEORETICAL ANALYSIS

Random Forest

Random forest is a commonly used tool in the construction of Decision tress. 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.
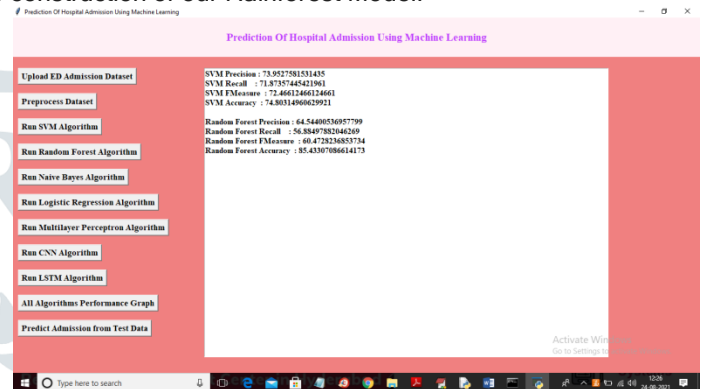

Fig:ii

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
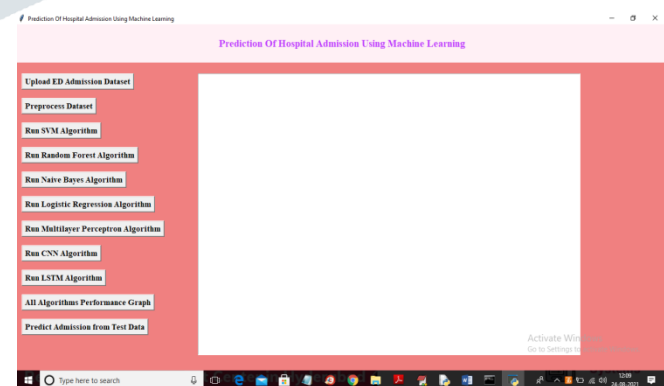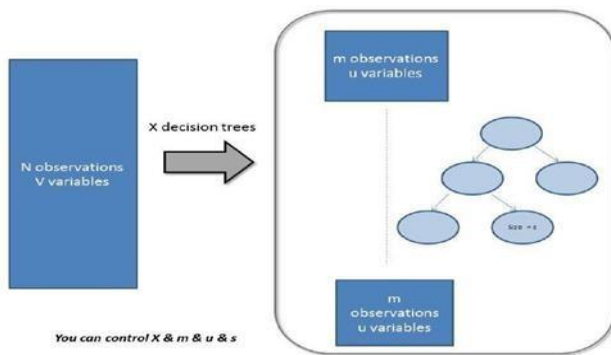

Fig:iii

Fig:iv
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 correlational 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 towards specific features. However as long as your able to keep these drawbacks in mind and have them rectified further down the line there shouldn't be any problem with using these libraries on your data.

PSEUDO CODE FOR RANDOM FOREST
1. Assume that N is the no of cases in our training set. Once you've done these take a sample amongst these N cases at random and with replacement.
2. Take the number of input features or 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 mode. M is then further used in order to split the nodes and known as best split. As we further construct our forest the value m will remain 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 Major concept behind the Random Forest Model

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 they are more trees in the model, the algorithm would not over fit 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 into their previous records.
4. It can be used in order to identify the behaviour of the stock market.
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 apply 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 con 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 has included areas such as hand written 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.
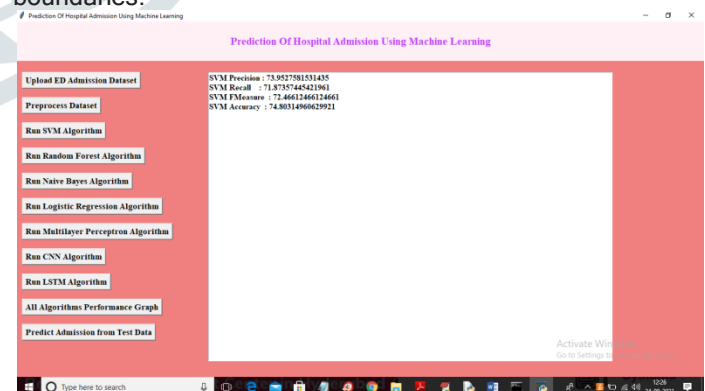

Fig:v

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 require far 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 then 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 hyperplane classifier and SVM's are the best models in order to handle these types of classifications and tasks.
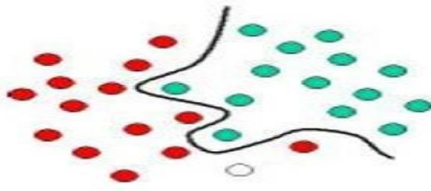

Fig:vi

The SVM's are able to take a complex hyperplane classification as above and turn it into a simple Linear Classification such as our first example. Let's take a closer look into how the SVM is able to do this with the diagram given below. In the diagram we can see how the original objects that have been classified are being rearranged using various mathematical formulae's. This process of rearrangement is known as mapping in the SVM. Once the SVM is able to map all of the objects these objects can be easily classified with the help of linear classification instead of having to use a complex hyperplane classifier. In order to establish this the SVM just has to find an optimal line which will be able to easily separate the two different objects and map all of them according to this line.
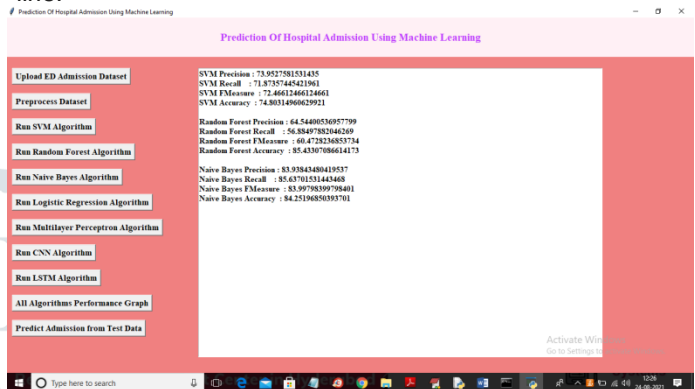

Fig:vii

Naïve Bayes:

Naive Bayes is a form of classification that is based upon Bayes Theorem which assumes that each and every predictor is independent. To phrase that in simple terms it means that Naïve Bayes assumes that a certain feature which is present in a class is completely independent and un relatable to all the other features that are also present in the same class. For example, let's take the case of a fruit being classified as an orange. A fruit is considered to be an orange if it is orange in color, round in shape, and roughly 3 inches in diameter. All these features combined are what classify a fruit to be an orange. If we are able to take these features independent from one another to calculate the possibilities of a fruit being an orange, then such form of individual prediction would be termed as _Naïve'. It is rather easy to construct a Naïve Bayes model and it is rather useful when we must deal with large datasets. Not only is this a fairly simple model to create it can easily outperform various other models which use highly sophisticated methods of classification. We are able to calculate the posterior probability P(a|y) from Pnaïve, P(y) and P(a|y) with the help of Bayes theorem.

Working Principle of Naïve Bayes Algorithm:

STEPS:
1. Transform all the data that we have into frequency tables.
2. Calculate all of the probabilities and create a likelihood table with them.
3. Using Naïve Bayes equation calculate the posterior probability for each and every class. The maximum

Advantages of Naïve Bayes:
1. It is rather simple and very quick in order to predict the

class of test data sets. It also performs exceptionally well when handling multi class predictions.
2. It is able to perform rather better than other models such as logistic regressions and it also requires far less training data.
3. Compared to numerical variables Naïve Bayes performs far better with categorical input variables.

Disadvantages of Naïve Bayes:
1. If the model does not observe a category of a categorical variable present within the training data set, the model will be unable to make a prediction and will label it as _0' probability. This is known as ¯Zero Frequency‖. However, this issue can be overcome by using various smoothing techniques.
1. It is also widely known as a bad estimator. Due to this the probability outputs that are taken from the predict probability are not taken seriously.
2. The major drawback of Naïve Bayes is how it assumes each predictor to be independent from one another. Such independency is merely impossible in real life.

Applications of Naïve Bayes:
1. Real Time Prediction: It is a very eager and fast learning algorithm. Due to this reason it is often implemented to create real time prediction models.
2. Multi-Class Prediction: As we have talked about before Naïve Bayes is also well known for its ability to make multiclass predictions. Here we can accurately predict the multiple classes of our target object.
3. Text classification/ Spam Filtering/ Sentiment Analysis: Naïve Bayes Classifiers are commonly used in text

4. Recommendation System: When Naïve Bayes Classifier and Collaborative Filtering work along with one another it is known as a Recommendation System. Recommendations Systems use ML as well as data mining techniques in order to filter all of the unseen information and use them to predict weather the user would enjoy the resource or hand.

Further Improvements of Naïve Bayes:

1. If the continuous features are not in a normal distribution, then we must various methods or transform it into a normal distribution.

2. If we experience a zero frequency in our test data, then we must implement soothing techniques such as ¯Laplace Correction‖ in order to make further predictions.

3. We should remove all the correlated features in advance as they tend to be voted twice in our model which leads to inflated importance.

4. Naïve Bayes Classifiers have a very limited number of selections when it comes to parameter tuning such as alpha=1 for soothing in order to gain knowledge upon the class's prior probabilities. This is why we recommend that we focus upon pre-processing the data and feature selection.

5. We have read various articles where people suggested to implement classifier combinations techniques such as bagging, ensembling and boosting. However, we believe that it would simply be a waste of time as their purpose is to mainly reduce the amount of variance. It is clearly visible that Naïve Bayes does not show any variance.

3. DATA MINING:

When we sort through large data sets in order to identify various patters 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.

CNN Algorithm:

In the past few decades, Deep Learning has proved to be a very powerful tool because of its ability to handle large amounts of data. The interest to use hidden layers has surpassed traditional techniques, especially in pattern recognition. One of the most popular deep neural networks is Convolutional Neural Networks (also known as CNN or ConvNet) in deep learning, especially when it comes to Computer Vision applications.

Since the 1950s, the early days of AI, researchers have struggled to make a system that can understand visual data. In the following years, this field came to be known as Computer Vision. In 2012, computer vision took a quantum leap when a group of researchers from the University of Toronto developed an AI model that surpassed the best image recognition algorithms, and that too by a large margin.

The AI system, which became known as AlexNet (named after its main creator, Alex Krizhevsky), won the 2012 ImageNet computer vision contest with an amazing 85 percent accuracy. The runner-up scored a modest 74 percent on the test. Convolutional neural networks are composed of multiple layers of artificial neurons. Artificial neurons, a rough imitation of their biological counterparts, are mathematical functions that calculate the weighted sum of multiple inputs and outputs an activation value. When you input an image in a ConvNet, each layer generates several activation functions that are passed on to the next layer.

The first layer usually extracts basic features such as horizontal or diagonal edges. This output is passed on to the next layer which detects more complex features such as corners or combinational edges. As we move deeper into the network it can identify even more complex features such as objects, faces, etc.
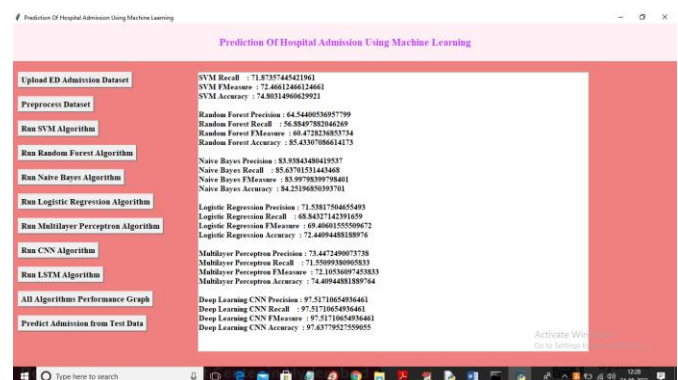

Fig:viii

LSTM Algorithm:

Long Short-Term Memory Networks is a deep learning, sequential neural network that allows information to persist. It is a special type of Recurrent Neural Network which is capable of handling the vanishing gradient problem faced by RNN.LSTM was designed by Hochreiter and Schmidhuber that resolves the problem caused by traditional rnns and machine learning algorithms. LSTM can be implemented in Python using the Keras library.

Let's say while watching a video, you remember the previous scene, or while reading a book, you know what happened in the earlier chapter. RNNs work similarly; they remember the previous information and use it for processing the current input. The shortcoming of RNN is they cannot remember long-term dependencies due to vanishing gradient. LSTMs are explicitly designed to avoid long-term dependency problems.

This article will cover all the basics about LSTM, including its meaning, architecture, applications, and gates.

In the introduction to long short-term memory, we learned that it resolves the vanishing gradient problem faced by RNN, so now, in this section, we will see how it resolves this problem by learning the architecture of the LSTM. At a high level, LSTM works very much like an RNN cell. Here is the internal functioning of the LSTM network. The LSTM network architecture consists of three parts, as shown in the image below, and each part performs an individual function.
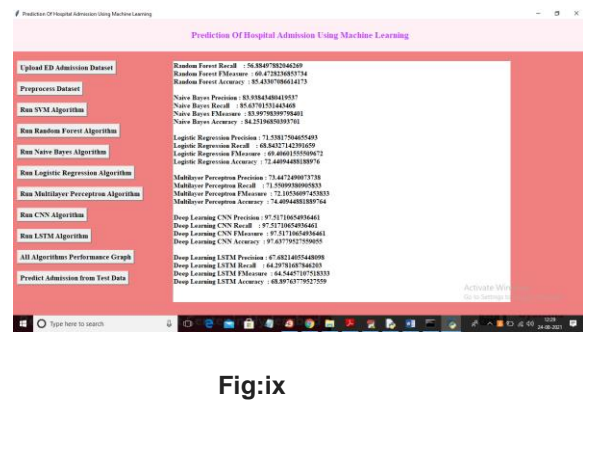

Fig:ix

3.OUTPUTS:

4.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 Naive 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.

5. 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. & Dorai Rangaswamy, 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.

Journal of Engineering and Technology(UAE), vol. 7, no. 2, pp. 68-71.

[4] Emergency Department Overcrowding in Massachusetts. Making Room in our Hospitals. Issue Brief.The Massachusetts Health Policy Forum, No 12; 2001.

[5] National Hospital Ambulatory Medical Care Survey. 2002 Emergency Department Summary. Advance Data Number 340.35pp.(PHS) 2004-1250.

[6] 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.

[7] K.S.S. Joseph Sastry & T. Gunashekar,' A Systematic Access Through Machine Learning Methods For Expectation In Malady Related Qualities', International Journal of Engineering and Advanced Technology (IJEAT), Volume-8, Issue-6S, August 2019, ISSN: 2249 – 8958

[8] 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

[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. Sripath Roy,K.Roopkanth, V.Uday Teja, V.Bhavana, & J.Priyanka,' Student Career Prediction Using Advaced Machine Learning Techniques',International Journal of Engineering & Technology, 7 (2.20) (2018) 26-29

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

[12] 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

[13] 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.