IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657
Volume 5,Issue 2,April 2017

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 helpsin 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 MedicalIndustry is Emergency Department Crowding. These are themost severely injured or patients who need immediateattention. However, it is often very difficult to identify the stateof all the patients in the Emergency ward which leads tomaking 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 overbut 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 toreact due to the shortage of required staff. This is why it is necessary for us to come up withinnovative approaches to solve this global issue to improve the patient flow and preventingpatient crowding.

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

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 5,Issue 2,April 2017

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 theemergencyward are transferred to these inpatient beds. This is one of theproblems we can easily rectify with the help of data mining inorder to identify patients that are inpatient admissions fromthose who are not so that we can avoid any confusions in oursystem. In thisstudy we will mainly focus uponimplementingvarious machine learning algorithms and developing models inorder to predict the state of the patients that are being admitted into the emergency department. We will also becomparing the performance of our model with a few variousapproaches that are already in the world. Patients who plan onvisiting the hospitals for various issues andthose that are inthe 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 uponthe 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.

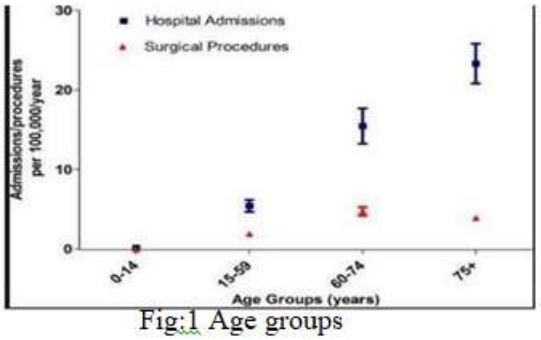


Fig:1 Age groups

The emergency attendees may come through main receptionor in ambulance at this point of time depending on thesituation of the patient the details should be taken for some of the medications age, gender, blood pressure, diabetes plays avital role for the further treatment. Usually, to collect the datafrom 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 suchcases, we must use a TriageScale in order to understand the condition of the patient andhow urgently they require medical care. This is one of themost important phases for the safety of any patient. When welook into the previous records of any hospital, we can clearlyidentify that there were far more aged people admitted in the hospital when compared to children or adults. This has causedchaos at emergency departments due to a lack of knowledgeBlockchain is the main toolto felicitate this need and when combined withdifferent hashing techniques, this becomes apowerful method for protecting the data. It alsohelps in eliminating the need for constantverification of certificates. Blockchain technologyis used to reduce the incidence of certificateforgeries 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, thestorage of certificates

ISSN 2347–3657
Volume 5,Issue 2,April 2017

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

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 createa quick and accurate Triage System in order to assess all the patients. Oncethe patient has undergone the Triage Processthey will shifted to the clinical room where they will be consulted by a clinician who willprovide the best course of action for the patient. There are various Triage Systems thatare used commonly around the world. However, the two most commonly used triage systems are those that use either a 3Level Triage System or a 5 Level Triage System. 3Level Triage System labels patients as Emergent, Urgent and Non urgent from the highest to lowest level respectively. Similarly, the5 Level System is broken down as Resuscitation, Emergent, Urgent, Less Urgent, Not Urgent from lowest levelto highest levelrespectively.

Various studies around theworldhave showed that the 5 Level TriageSystem has been farmore reliable than the standard 3 Level System. It has done abetter job in predicting the consumption of resources, length ofstay, admission rates and mortality. Building a Triage Systemthat is highly accurate and precise can play a major impact inthe medical industry as it could save millions oflives. Our

Study is based upon two major objectives. Our first objective isto create and develop a model that is able to accuratelypredict weather a patient from the emergency department willbe admitted into the hospital. Our Later objective is to studythe performance of various other machine learning algorithmsin this sector. In order to predict the state of a patient we mustfirst have our heads wrapped around the knowledge of variousmathematical models. The previous research was done byusing logistic regression, decision tree and time seriesforecasting algorithms. In the previous analysis whencompared with other algorithms like logistic regression,decision tree and gradient boosted. Gradient boosted got themoreaccurate as we use decision tree it is not suitable forlonger data sets and need to perform pruning in decision treeswhereas in gradient boosted it merges the weaker trees andforms the stronger one which helps in the prediction.According to the statistics the rate of patient stays, or visitswas gradually increased from the year 2005 to 2014. Annual average growth rate for impatient stay was 5.7% andcumulative increase was 64.1% where as in EmergencyDepartment visits annual average growth rate was 8.0% andcumulative increase was 99.4%. Objective is to find the modelwhich suits the best and gives the accurate results forpredicting the admission in the emergency department. Here the comparison of three machine learning algorithms wasdone (i.e.) Naïve Bayes, Support vector machine (SVM),Random Forest classifier. After comparison Support Vectormachine got the most accurate results when compared toothers.

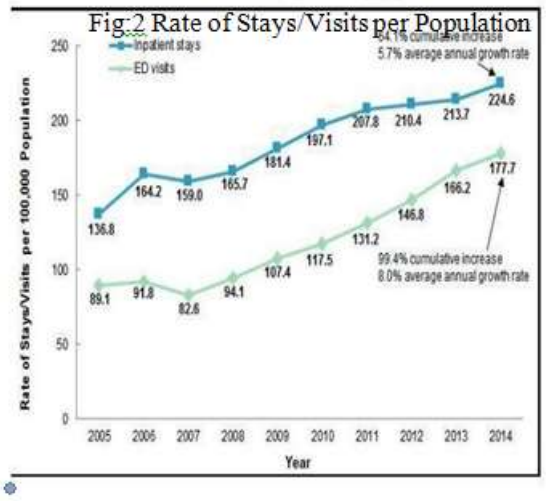


Fig:2 Rate of Stays/Visits per Population

2.LITERATURE REPORT

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657
Volume 5,Issue 2,April 2017

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: Internetin 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 foradoption to different models.

Impact of streaming "fast track" emergencydepartment 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.

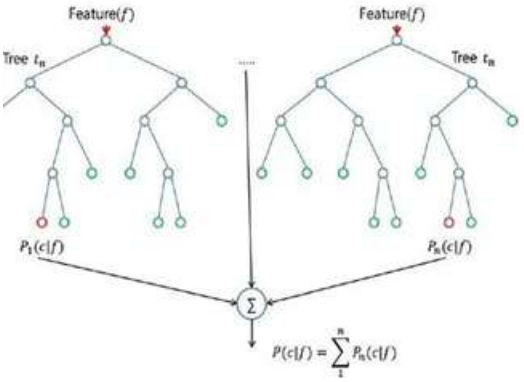
Methods: During the 12-week trial period, patients in Australasian Triage Scale (ATS) categories 3, 4 and 5 who were likely to be discharged were identified at triage and assessed and treated in a separate fast track area by ED medical and nursing staff rosteredto work exclusively in the area.

A Comparative Study on Feature Selectionin 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 Ø 2 -test (CHI), and term strength (TS). We found IGand CHI most effective in our experiments. Using IG thresholding with a knearest 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 thecomputation of these measures are too expensive. TS compares favorably with the other methods with up to 50% vocabulary redu.**emergency 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 forpatient 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 theimplementation of a process of seeing those patients that minimized complex queuing in theED.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 5,Issue 2,April 2017



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 ofDecision tress. Instead of following the normal routine it takesa subset of variables and observations in order to constructthe decision tree. It builds various decision trees and mergesthem together in order to form a single decision tree that hashighaccuracy and prediction. The Random Forest isgenerallyviewed upon as a black box as its predictions are highlyaccurate. Most people don ‘t bother about the backgroundcalculations due to its high accuracy rate. Although we won ‘tbe able to change the methods of calculations for the RandomForest, it has a few modifiable factors which can in turn effectthe performanceof the model or the resources and timebalance. We will talk about their variable factors further on inthe construction of our Rainforest Model.

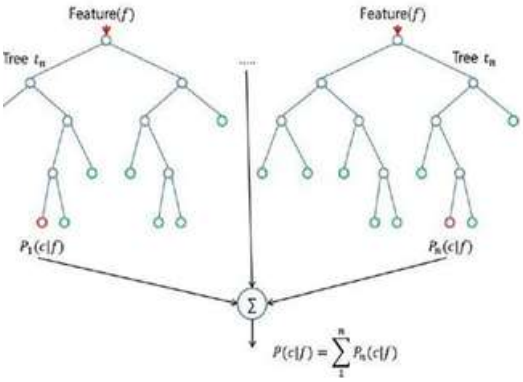


Fig:3 Random Forest

Parameters to Tune Random Forests

There are 2 major roles that the parameters play when weconstruct our decision tree. The parameters can either effectthe prediction power of our model or they allow us to trainourmodel far more simply. Let‘s take a look into these parametersin a far more detailedmanner

Major concept behind the

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 5,Issue 2,April 2017

Rrectified further down theline there shouldn't a be any problem with using these librarieson n your data.
d
o
m
F
o
r
e
s
t
M
o
d
e
l

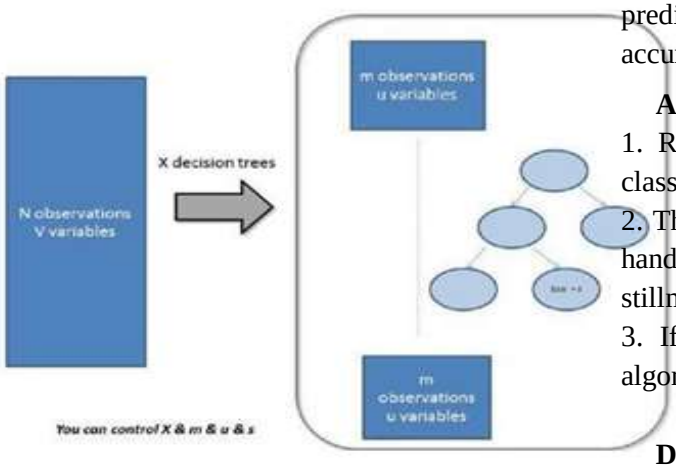


Fig:4

FIG:4 Features which make predictions to the model betterthe chance of bumping into impurities are relatively high whenit comes to using random rainforest libraries. However, they allcome with their drawbacks when you keep data interpretationin mind. A simple example for this is using correlationalanalysis. A feature that typically has a very strong score canbe depicted as a low score feature within rainforest. Anotherrecurring impurity is how certain methods are biased towardsspecific features. However as long as your able to keep thesedrawbacks in mind and have them

PSEUDO CODE FOR RANDOM FOREST

1. Assume that N is the no of cases in our training set. Onceyou 've done these take a sample amongst these N cases atrandom and with replacement.
2. Take the number of input features or variables as M. Wemust then specify a number n such that n<M and also thatthere are m input variables selected at each mode. M is thenfurther used in order to split the nodes and known as bestsplit. As we further construct ourforest the value m will remain constant
3. If pruning does not occur each tree is allowed to grow aslarge as possible.
4. All the constructed trees are merged together to form asingle tree to create predictions with much higher accuracy.Majority Voting is the

Advantages of Random Forest

1. Random Forest algorithm deals with both classification andregression tasks.
2. The Random Forest Model is able to handle any missing values in our data set and stillmaintain predictions with highaccuracy.
3. If they are more trees in the model, the algorithm would notover fit the model.

Disadvantages of Random Forest

1. Good at classification concept but not so good atRegression.
2. As the inner calculations of the Random Forest Model arescarcely known we have very little control on how the modelfunctions.

Applications

1. They are used within banking sectors in order to segregateloyal and fraud clients.
2. It is used within the medical industry in order to identify thepossible combinations of various components in order tovalidate the correct medicine.
3. It is further used in order to label patients with themrespective problems by looking

ISSN 2347–3657
Volume 5,Issue 2,April 2017

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 besimply 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 originaltraining data into training data in higher dimensions. Once wehave established this new dimension the model will beginsearching for linear optimal separating hyperplanes. The SVMis able to find and separate these hyperplanes with the usageof support vectors and margins. We will look deeper into theseconcepts later on in our study. However, in the past decadeSVM‘s have

been attracting a lot of attention. SVM‘s were firstintroduced into the picture when Vladimir Vapnik along with hiscolleagues Bernhard Boser& Isabelle Guy decided to write apaperon them in 1992. Although these group of researcherswere the first to have written a paper on SVM‘s the concepthas dated back to the 1960s. SVM‘s follow a rathercomplicated internal structure and the time to train them isextremely 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 tooverfitting. A commonly used application of SVM‘s has beennumeric or alphanumeric prediction as well as classification.Other applications for SVM‘s has included areas such as handwritten language or digit detection, speaker identification,object detection, and Benchmark time series. SVM‘s aremostly based upon the concepts of decision planes that havepredefined boundaries. A decision plane can simply bedefined as a barrier that separates the various objects thatbelong to different membership classes. Let‘s try to take a lookat this simple schematic example in which objects eitherbelong to the left class or the right class. The line in the middleacts as the boundary or you can say decision plane whichseparates the right and left class. All the objects thataresituated to the left of this line are known as the left class whileall those to the right areclassified as the right class. When anew object enters into the scenario it falls upon the boundaryline which will then make the classification to either push it leftor right into its respective class

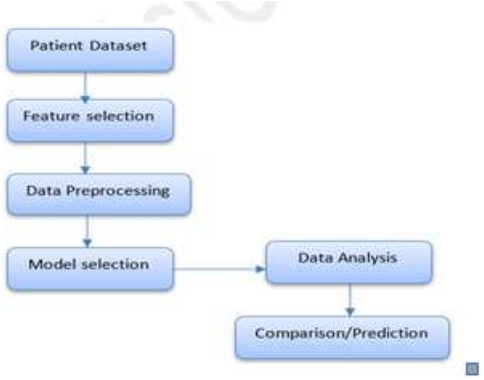


Fig: 5

The schematic example that we have lookedinto above isnothing more than a basic linear classifier. A linear classifier isnothing morethan a classifier that separates the objects into various groups. In our example these groupswere color basedof Red & Green objects. Thisis a very basic and simplemethod of classification. However not all classifications are assimple as this one, they are often far morecomplex require farmore classifications inorder to properly segregate the trainingobjects.Let us take a look at another classification withthesame segregation of Red & Green objectsbelow.

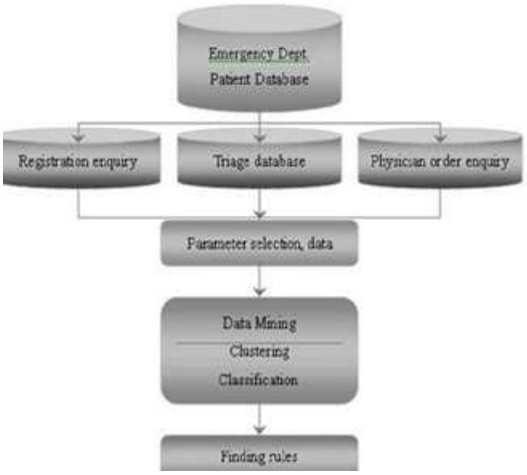
Comparedto our previous linear classification we can clearly see that

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 5,Issue 2,April 2017

theseparation of the objects now requires a curve then a line. Acurve is a far more complexstructure then a line. Theclassification for thiscurve takes place by drawing variousseparatinglines in order to identify and distinguish theobjectsfrom one another. This type ofclassification is commonlyknown ashyperplane classifier and SVM's are thebestmodels in order to handle these types ofclassifications andtasks.

4. DATA MINING

When we sort through large data sets in order to identifyvarious patters and establish relationships it is known asData Mining. These patterns and Relationships can befurther used in order to solve various problems through dataanalytics. Enterprises are able to make predictions uponfuture trends with the help of Data Mining tools. We are ableto do so by using massive amounts of data in order toidentify the various patterns and trends. It typically consistsof Data Transformation, Pattern Evaluation, Data Cleaning,Pattern Discovery, Data Integration. andKnowledgePresentation. We use Association rules within data miningby exploring andanalyzing the data for various if/thenpatterns. From here we will use various support andconfidence criteria in order to form various importantrelationships among data. Support is defined as the numberof times a specific query is found within a database, whileconfidence is the probability that the if/then case isaccurate. There are other parameters used within datamining such as Sequence or Path Analysis, Clustering andForecasting, Classification, and Sequence or Path Analysis.An ordered list of aset of items is known as a Sequence. Itiscommonly found in any sort of Database. TheClassification Parameter is used in order to detect newpatterns, It may also change thestructure of our organizeddata. All Data Mining techniques are executed in a specific organized manner or flow. For you to get abetterunderstanding have a look at theflowchart 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. NaiveBayes 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

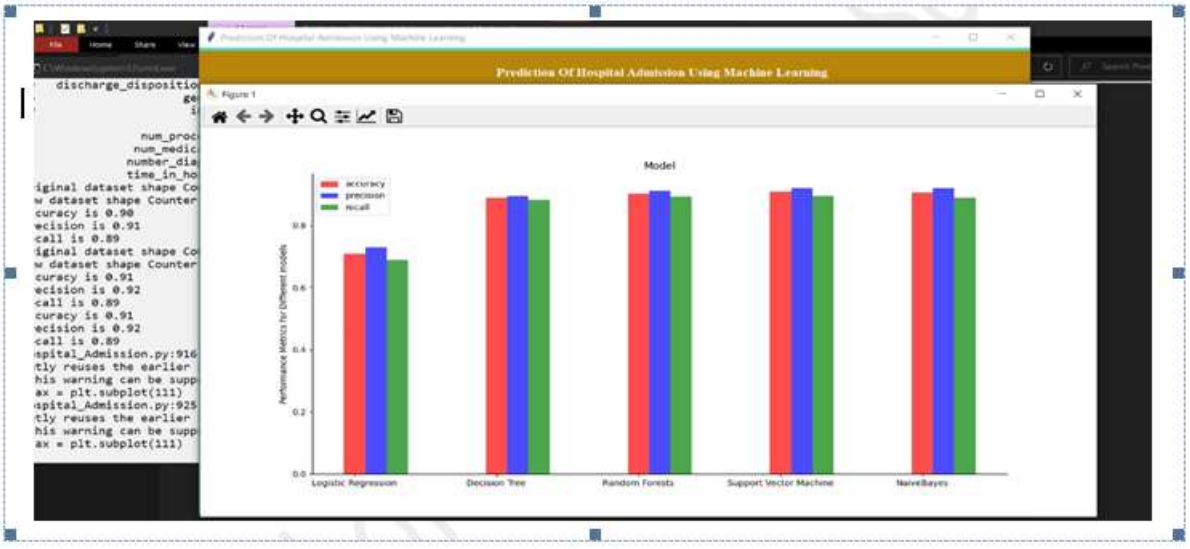
In above screen home page for Patient

ISSN 2347–3657
Volume 5,Issue 2,April 2017

Admission Prediction, In the above Screen Patient Data set wasloaded.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 5,Issue 2,April 2017

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



In above screen it displaying about Model comparison

ISSN 2347–3657
Volume 5,Issue 2,April 2017

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 thatwe 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 modelsthat 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 inthe 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 uponthese 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, BlondellKet al. Impact of streaming _fast track' emergency department patients. Aust.Health Rev. 2006; 30: 525–32.

[4] King DL, Ben-Tovim DI, BasshamJ.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 incurrent Indian web documents", International [5] L. Yi, B. Liu, and X. Li, 'Eliminating Noisy Information in Web Pages for Data Mining', in Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2003, pp. 296–305. 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 Number340.35pp.(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

ISSN 2347–3657
Volume 5,Issue 2,April 2017

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

[10] 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, 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 Advaced 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.