A Data Driven Approach to Predict Emergency Admissions in Healthcare Units

[1]Vennamaneni Reshma, [2]Manoj Kumar Gottimukkala , [3]M.Jayapal

Student, Department of Computer Science and Engineering,Malla Reddy college of Engineering and Technology, Hyderabad, Telangana. India

Assistant Professor, Department of Computer Science and Engineering,Malla Reddy college of Engineering and Technology, Hyderabad, Telangana. India

Associate Professor, Department of Computer Science and Engineering,Malla Reddy college of Engineering and Technology, Hyderabad, Telangana. India

Abstract

Emergency department (ED) plays an important role in healthcare industry. It is the first department in case of personal health emergency to give soothing support services to patients. Afterwards, the patient may be shifted to corresponding department based on the advice of ED. Due to various reasons, there is increasing number of admissions in ED. In such cases, there needs to be a technology driven approach that can help in analysing situations and take appropriate decisions. This will help to regularize the flow of patients and avoid overcrowding. The data analysis or data mining offline can help to gain correct Business Intelligence (BI) or actionable knowledge. Machine learning algorithms that are part of Artificial Intelligence (AI) can solve complex problems as well. There are many machine learning approaches. In the existing system considered, machine leering algorithms like Decision Trees (DT), Gradient Boosted Machines (GBM) and Logistic Regression (LG) are employed for gaining knowledge. Different performance metrics are used to evaluate them. However, it is understood that the algorithms do not have support for feature selection. Unless feature selection algorithm is used, the classification algorithms will show mediocre performance due to redundancy in data and also presence of irrelevant features. Therefore, in this paper we develop a framework with some more prediction models equipped with feature selection knowledge to enhance the state of the art.

Keywords: Hospital admissions prediction, data mining, machine learning, feature selection

1. INTRODUCTION

In health care domain, it is essential to have improvisation of various scenarios. One such scenario is that emergency admissions. There are many researchers focused on this scenario and used machine learning (ML) algorithms for the same. As machine learning helps in understanding the trends in the data and make predictions, supervised learning methods are widely used. In this paper different supervised learning methods are used along with a hybrid feature selection method. Unless feature selection algorithm is used, the classification algorithms will show mediocre performance due to redundancy in data and also presence of irrelevant features. The proposed feature selection algorithm improves performance of the prediction models. It is thus able to predict emergency admissions in a hospital well so as to help administrators to envisage the demand and take necessary actions. Our contributions in this paper are as follows.

1. We proposed a framework based on machine learning (ML) algorithms used as prediction models for emergency admissions in a hospital.
2. We proposed a feature selection algorithm for improving performance of prediction models with higher quality in training phase. The algorithm is known as Hybrid Feature Selection (HPS).
3. A prototype is built to demonstrate proof of the concept. The application is used to evaluate the proposed framework in the prediction of emergency admissions.

2. RELATED WORK

This section reviews literature related to the proposed system. As explored in [1], emergency department in hospitals overcrowding. They also analysed the developed a logistic regression model to predict the probability of admissions at triage, using two years of routine administration data collected from hospitals in Glasgow. The most important predictors in their model included `triage category, age, National Early Warning Score, arrival by ambulance, referral source, and admission within the last year', with an area under the curve of the receiver operating characteristic. Boyle et al. [2] used routine administrative data to predict emergency admissions, also using a logistic regression model. However, their model was less accurate with an accuracy of 76% for their best model.

Fatovich et al. [3] achieved better performance using a Coxian Phase model over logistic regression model, with the former AUC-ROC of 0.89, and the latter 0.83. McCarthy et al. [4] developed three models to predict ED admissions using logistic regression models, naïve Bayes, and expert opinion. All three techniques were useful in predicting ED admissions. Variables in the model included age, arrival mode, emergency severity index, designation, primary complaint, and ED provider. Their logistic regression model was the most accurate in predicting ED admissions, with an AUC-ROC of 0.887. Perhaps surprisingly, this model performed better than triage nurse's opinion regarding likely admission. As there are many approaches for prediction of hospital admissions. The researches include [5]-[20] has different approaches that are used to model the runtime situations in the hospitals and the admissions are understood in terms of trends. From the literature it is understood that there is need for more comprehensive framework to get desired business intelligence.

3. PROPOSED METHODOLOGY

The proposed methodology for prediction of hospital admissions is provided here. The methodology is based on machine learning algorithms. Machine learning algorithms have been around and they perform well when training data is available. As the classification is based on the training set. There is need for quality in training of data. Towards this end, a feature selection algorithm known as “Hybrid Feature Selection” is achieved. The classification algorithms used for the empirical study include kNN, NB, DT, RF and SVM.

3.1 The Framework

The framework takes hospital admissions dataset which is subjected to data pre-processing. The data is divided into 80% training and 20% testing. The training set is given for feature selection. After feature selection, the features obtained by the algorithm are given to a classifier. The result of training the learned classifier or prediction model that provides classification results of test data.

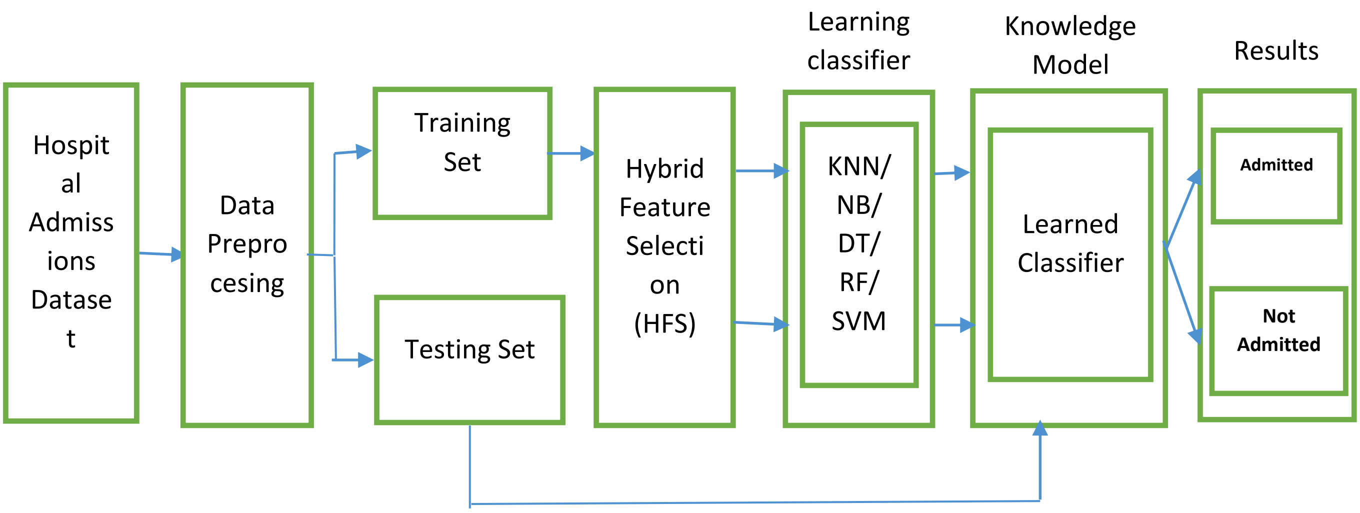

Figure 1: The proposed methodology

The framework used different machine learning algorithms. Each algorithm follows different approach in making predictions. However, they all are supervised learning methods that needs training prior to making predictions.

3.2 Prediction Models Used

Support Vector Machine (SVM) is the classification algorithm that creates optimal hyperplane that will have maximal margin. The margin pertaining to optimal hyperplane helps in discrimination of class lables. With the hyperplane creation, the SVM performs prediction of labels for unlabelled instances.

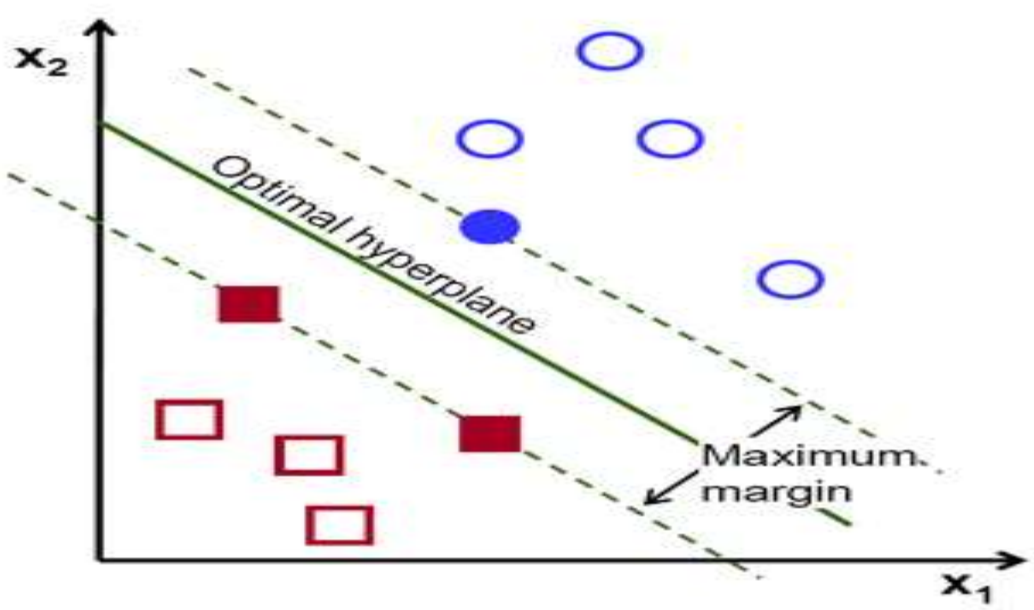

Figure 2: Functionality of support vector machine

Naïve Bayes is another classification algorithm used in the empirical study. This model depends on probability. It is widely used in different applications for prediction of class labels. NB is based on Bayes theory. It uses the naïve and independent features. It works faster because it is based on probabililty. K-NN is the classification algorithm that works without the knowledge of data priori. It finds k-nearest data points.

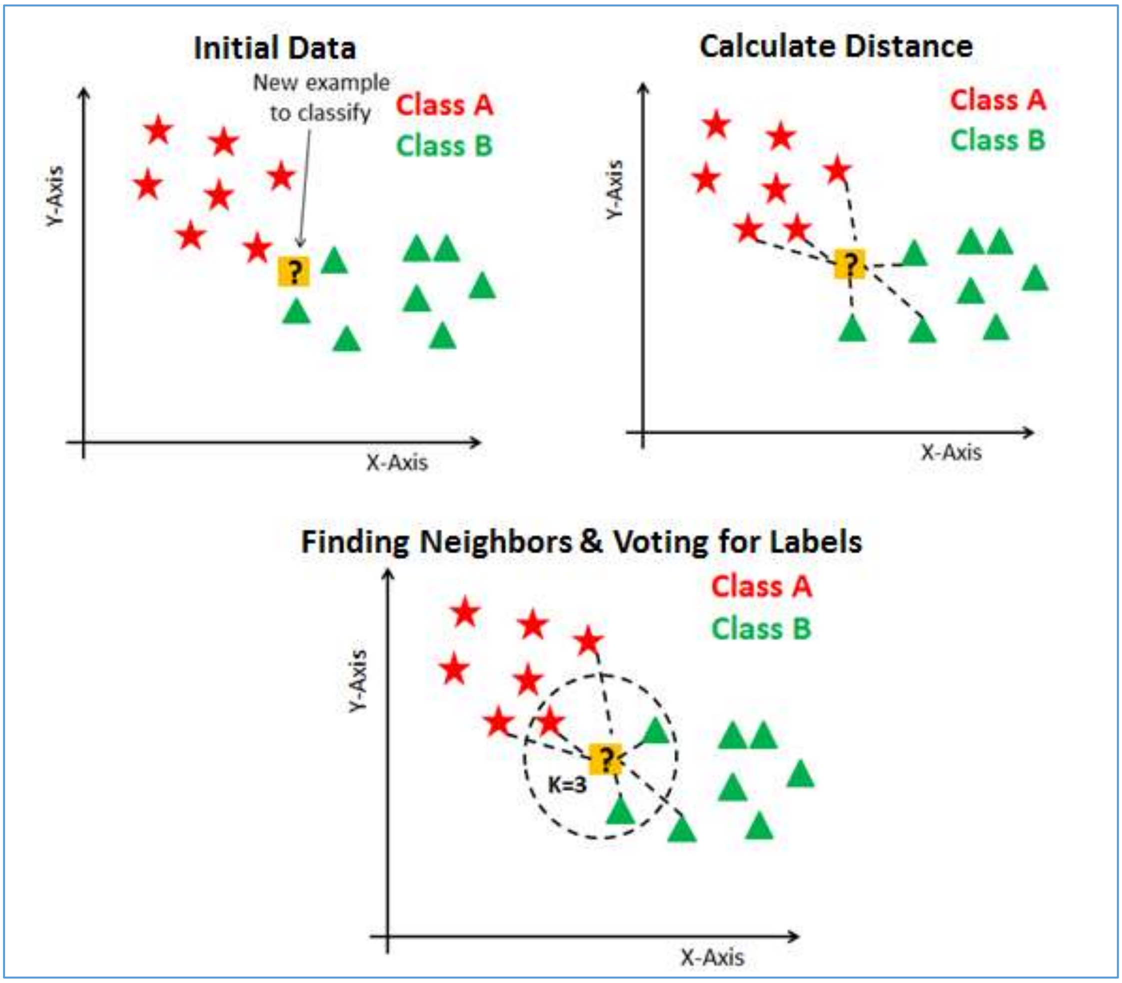

Figure 3: Process of kNN classification

As presented in Figure 3, the KNN algorithm has three phases. In the first phase, distance is computed from a point specified. In the second phase, neighbors which are closese are found and voted for labels. The data points that get more votes will be determined as class labels. Decision Tree (DT) is nother widely used classifier. It is a powerful tool that is used to predict labels towards classification. The restuls of DT can be understood by humans for making well informed decisions.

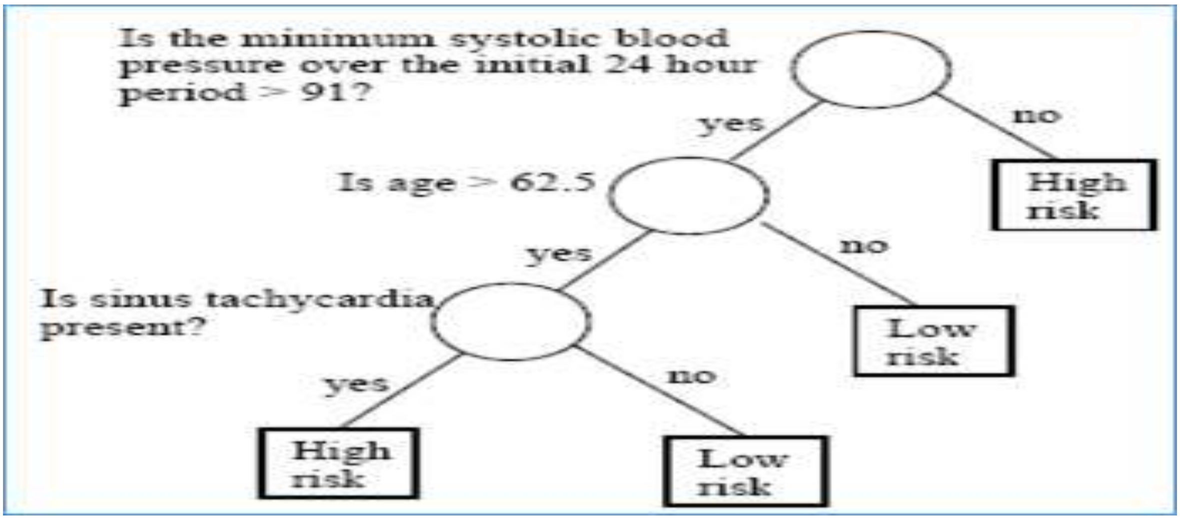

Figure 4: Process of decision tree with healthcare data

As presented in Figure 4, the healthcare data is subjected to classification. The resultant tree shows the decisions. Every condition in the decision rules contain two possible values known as yes or no. It works well for continuous and categorical data. The data is divided into multiple parts. Every atribute's entropy is computed. The attributes exhibiting maximum information gain and minimum entropy are selected to split data for providing decisions. Eq. 1 and Eq. 2 provides formulae for entropy and gain.

$$Entropy(S) = \sum_{i=1}^{c} -p_i \log_2 p_i \quad (1)$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

Yet another algorithm used for empirical study is known as Random Forest (RF). It is made up of multiple decision trees. Each decision tree is learned. When multiple decision trees are combined, it results in random forest.

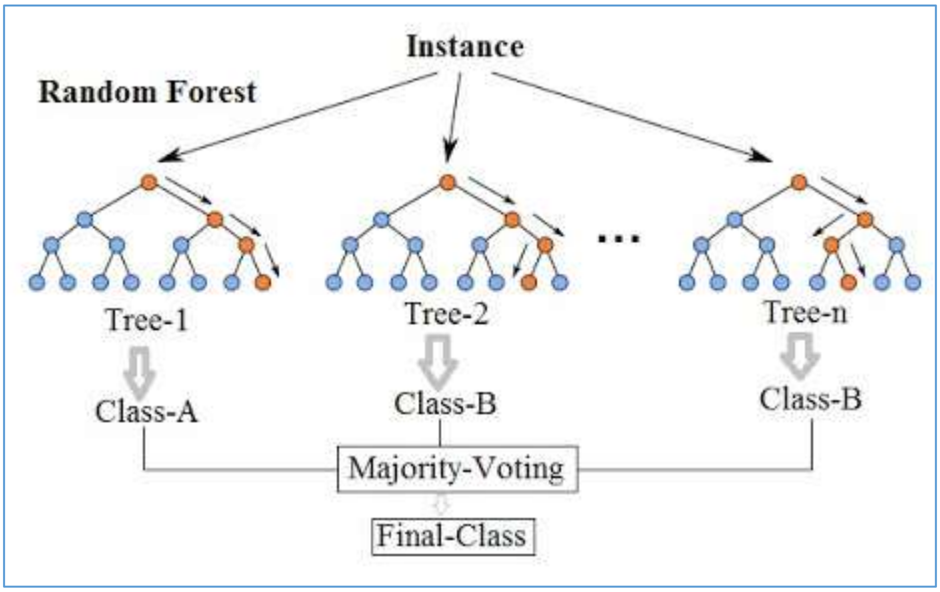

Figure 5: Process of Random Forest classification

As presented in Figure 5, the RF has many decision trees and based on the voting concept the final class labels are determined. The underlying process has a stop condition that is verified for each tree. Splitting the data is continued until a stop condition is met. RF can handle data with higher efficiency. Its process of estimating errors is unbiased. As it works on missing data, its accuracy is generally high.

3.3 Proposed Feature Selection Algorithm

The proposed feature selection algorithm is known as Hyrbid Feature Selection (HFS). It takes hospital adminssions training dataset and returns the features that are able to contribute to class label prediction.

Algorithm: Hybrid Feature Selection

Inputs: Hospital Admissions Training Set D

Output: Selected Features SF

```
Initialize attribute vector A
Initialize a tree T
A=LoadData(D)
For each attribute a in A
   e=findEntropy(a)
   g=findGain(a)
   IF e and g satisfy threshold THEN
      F=add(e)
   END IF
```

```
End For
For each feature in F
    T=add(feature)
End For
SF=findCorrelatedFeatures(T)
Return SF
```

Algorithm 1: Hybrid feature selection algorithm

As presented in Algorithm 1, the HFS algorithm uses two important measures like entropy and gain. These measures are used to know the features that are useful in prediction of classes. The correlation of features found in the process help in removing some of the features that are not able to help in prediction. Thus it results in dimentionality reduction. When the number of dimensions is reduced, it leads to quality in training and also helps in performance enhancement. The quality of training is thus improved. It paves way for efficient prediction of admisisons in emergency department of the hospital.

4. EXPERIMENTAL SETUP

Experiments are made with a prototype application developed using Java programming language. It has a web based interface to support prediction of hospital admissions based on the given dataset. The dataset contains around 100000 instance with around 9.2% missing data. The data is used for empirical study. Out of all instances 80% are used for training and 20% are used for testing. Performance evaluations made using the metrics based on the confusion matrix presented in Table 2.

	Ground Truth (correct prediction)	Ground Truth (incorrect prediction)
Result of algorithm (correct prediction)	True Positive (TP)	False Positive (FP)
Result of algorithm (incorrect prediction)	False Negative (FN)	True Negative (TN)

Table 1: Confusion matrix

Different metrics are used to evaluate the performance of the algorithms. They are known as precision, recall and F-measure as shown in Eq. (1), Eq. (2) and Eq. (3).

$$precision = \frac{TP}{TP + FP} \quad (1)$$

$$recall = \frac{TP}{TP + FP} \quad (2)$$

$$F - measure = \frac{2 \times precision \times recall}{precision + recall} \quad (3)$$

These measures help in evaluating the framework and all the classification algorithms. The harominic mean of precision and recall is known as F-Measure.

5. EXPERIMENTAL RESULTS

Experiments are made with the proposed methodology. The algorithms are executed with and without Hybrid Feature Selection defined. With feature selection, the quality of the training is imporved and the performance of the algorithm is improved in predicting the admissions in hospital.

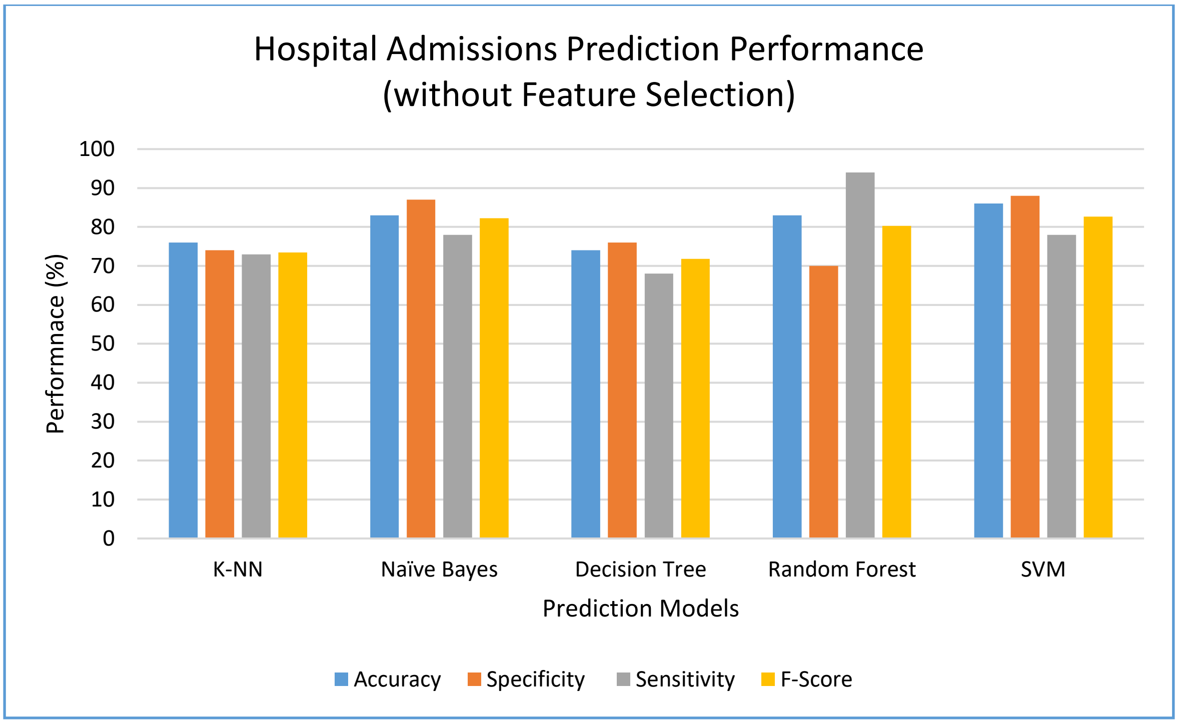

Figure 6: Performance comparision without feature selection

As presented in Figure 6, the machine learning algoirhtms (hospital admissions prediction models) like kNN, NB, DT, RF and SVM are provided in horizontal axis. The performance in terms of accuracy, specificity, sensitivity and F-score is presented in vertical axis. The results revealed that different machine learning algorithhms have different performance. The results are observed without feature selection.

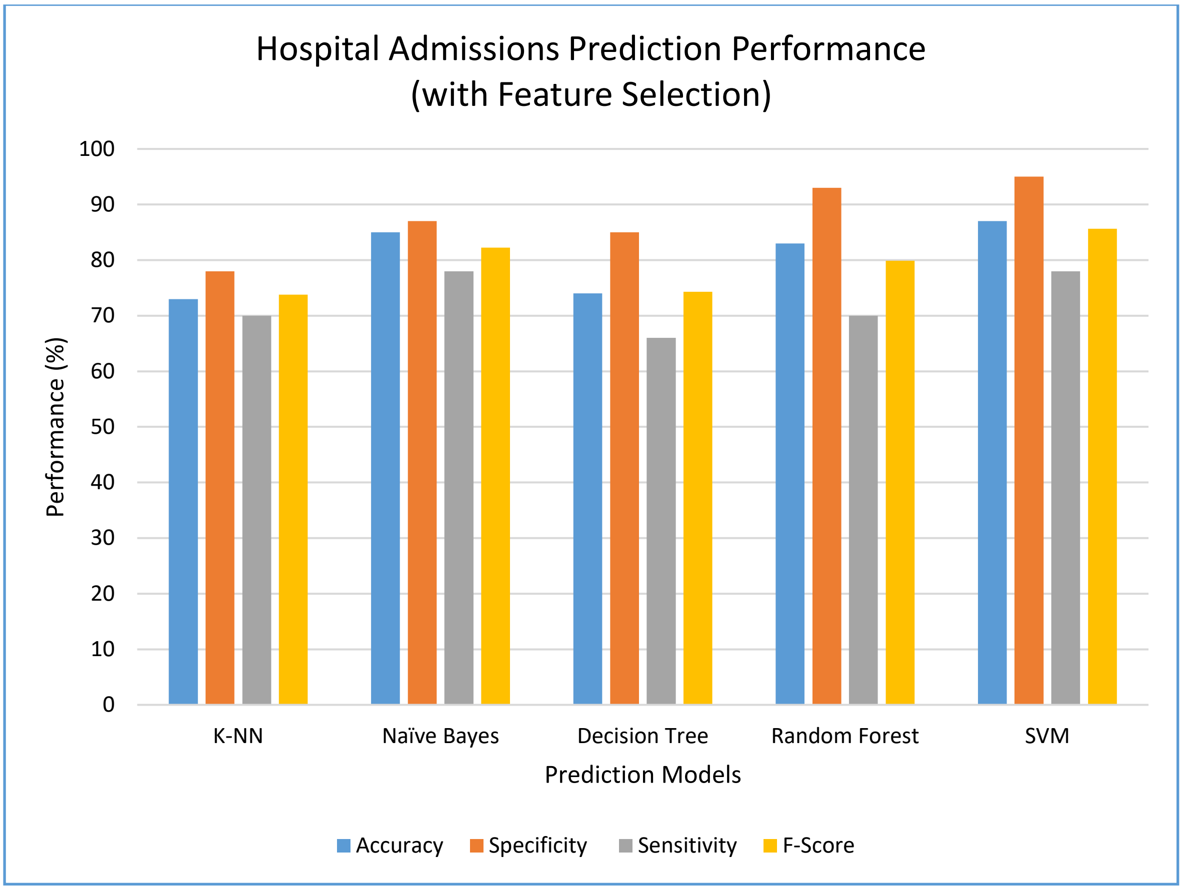

Figure 7: Performance comparison with feature selection

As presented in Figure 7, the machine learning algoirhtms (hospital admissions prediction models) like kNN, NB, DT, RF and SVM are provided in horizontal axis. The performance in terms of accuracy, specificity, sensitivity and F-score is presented in vertical axis. The results revealed that different machine learning algorithhms have different performance. The results are observed with feature selection. With feature selection algorihtm propsoed in this paper (Hybrid Feature Selection), the quality of the trining is improved. The feature selection causes the performance to be improved. When compared with the results of Figure 6, the results with feature selection shows better performance.

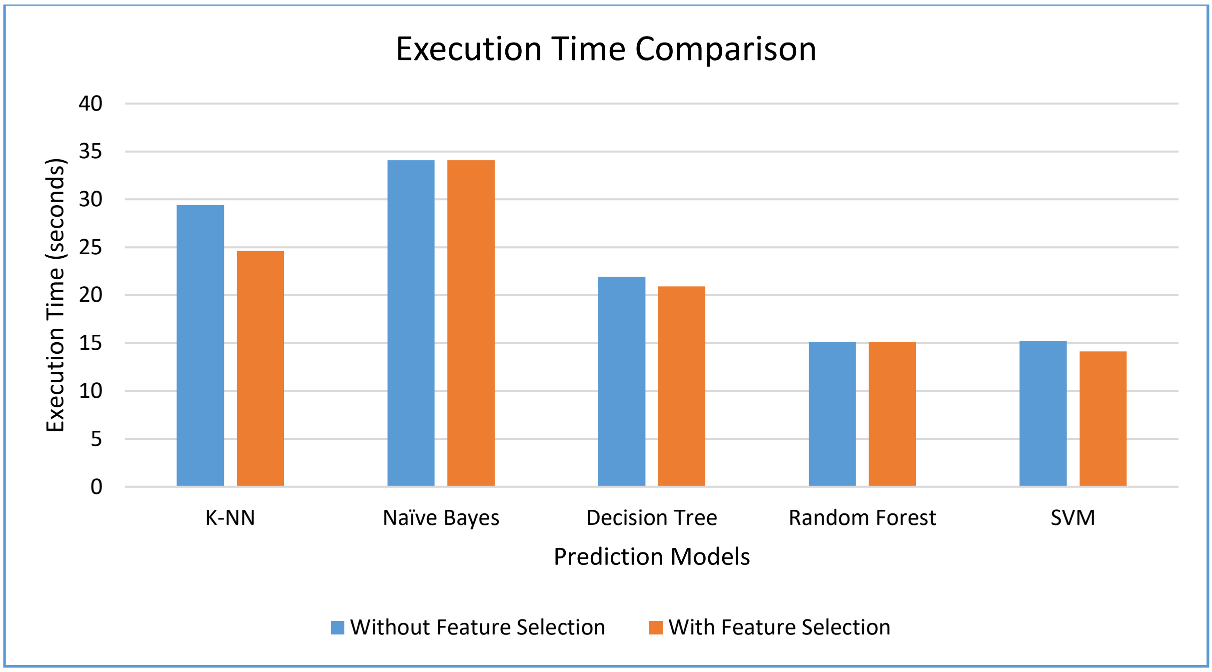

Figure 8: Execution time comparison

As presented in Figure 8, the horizontal axis shows the prediction models while the vertical axis shows the execution time in seconds. The results revealed that the prediction models showed better performance in terms of execution time. This is consistenly true with respect to all the models like KNN, NB, DT, RF and SVM. When features are selected, it reduces the time complexity. In other words, there is no need for processing unnecessary fields that do not contribute in the selection of class label.

6. CONCLUSIONS AND FUTURE WORK

Healthcare units are part of the healthcare industry where emergency departments have pressure of taking new adminssions and dealing with patients who need treatment from different departments. In such cases, the administrators are not aware of the trends in the admissions. To overcome this problem and help administrators to predict the admisions, in this paper, we proposed a ML based framework that leverages prediction process. Moreover, the proposed system has multiple classifiers that are supported by feature selection algorithm known as Hybrid Feature Selection (HBS) that results in selecting features from training set that are useful in prediction of class labels. The proposed feature selection algorithm improves performance of the prediction models. It is thus able to predict emergency admissions in a hospital well so as to help administrators to envisage the demand and take necessary actions. The experimental results revealed that the proposed feature selection algorithm improves performance of the prediction models. In future, we intend to use deep learning models for more accurate prediction of emergency admissions in the hospital.

References

[1] J. S. Olshaker and N. K. Rathlev, ‘‘Emergency department overcrowding and ambulance diversion: The impact and potential solutions of extended boarding of admitted patients in the emergency department,’’ J. Emerg. Med., vol. 30, pp. 351–356, Apr. 2006, doi: 10.1016/j.jemermed.2005.05.023.

[2] J. Boyle et al., ‘‘Predicting emergency department admissions,’’ Emerg. Med. J., vol. 29, pp. 358–365, May 2012, doi: 10.1136/emj.2010.103531.

[3] S. L. Bernstein et al., ‘‘The effect of emergency department crowding on clinicallyorientedoutcomes,’’Acad.Emerg.Med.,vol.16,no.1,pp.1–10, 2009, doi: 10.1111/j.1553-2712.2008.00295.x.

[4] D. M. Fatovich, Y. Nagree, and P. Sprivulis, ‘‘Access block causes emergency department overcrowding and ambulance diversion in Perth, Western Australia,’’ Emerg. Med. J., vol. 22, no. 5, pp. 351–354, 2005, doi: 10.1136/emj.2004.018002.

[5] M. L. McCarthy et al., ‘‘Crowding delays treatment and lengthens emergency department length of stay, even among high-acuity patients,’’ Ann. Emerg. Med., vol. 54, no. 4, pp. 492–503, 2009, doi: 10.1016/j.annemergmed.2009.03.006.

[6] D. B. Richardson, ‘‘Increase in patient mortality at 10 days associated with emergency department overcrowding,’’ Med. J. Aust., vol. 184, no. 5, pp. 213–216, 2006.

[7] N.R.HootandD.Aronsky,‘‘Systematicreviewofemergencydepartment crowding:Causes,effects,andsolutions,’’Ann.Emerg.Med.,vol.52,no.2, pp. 126–136, 2008, doi: 10.1016/j.annemergmed.2008.03.014.

[8] Y.Sun,B.H.Heng,S.Y.Tay,andE.Seow,‘‘Predictinghospitaladmissions at emergency department triage using routine administrative data,’’ Acad. Emerg. Med., vol. 18, no. 8, pp. 844–850, 2011, doi: 10.1111/j.15532712.2011.01125.x.

[9] M. A. LaMantia et al., ‘‘Predicting hospital admission and returns to the emergency department for elderly patients,’’ Acad. Emerg. Med., vol. 17, no. 3, pp. 252–259, 2010, doi: 10.1111/j.1553-2712.2009.00675.x.

[10] J. S. Peck et al., ‘‘Generalizability of a simple approach for predicting hospital admission from an emergency department,’’ Acad. Emerg. Med., vol. 20, pp. 1156–1163, Nov. 2013, doi: 10.1111/acem.12244.

[11] A. Cameron, K. Rodgers, A. Ireland, R. Jamdar, and G. A. McKay, ‘‘A simpletool topredict admissionat thetime oftriage,’’ Emerg. Med. J., vol. 32, no. 3, pp. 174–179, 2013, doi: 10.1136/emermed-2013-203200.

[12] N. Esfandiari, M. R. Babavalian, A. M. E. Moghadam, and V. K. Tabar, ‘‘Knowledge discovery in medicine: Current issue and future trend,’’ Expert Syst. Appl., vol. 41, no. 9, pp. 4434–4463, 2014, doi: 10.1016/j.eswa.2014.01.011.

[13] H. C. Koh and G. Tan, ‘‘Data mining applications in healthcare,’’ J. Healthcare Inf. Manag., vol. 19, no. 2, pp. 64–72, 2005.

[14] C. Baker. (2015). Accident and Emergency Statistics, England. [Online]. Available: www.parliament.uk/briefing-papers/sn06964.pdf

[15] DHSSPS. (2015). Northern Ireland Hospital Statistics: Emergency Care (2014/15), Northern Ireland. [Online]. Available: https://www.dhsspsni.gov.uk/sites/default/files/publications/dhssps/hsemergency-care-2014-15.pdf

[16] M.-L. Connolly. (2015). NI Emergency Healthcare Enquiry Finds ‘Degrading’ Cases. Accessed: Oct. 14, 2015. [Online]. Available: http:// www.bbc.co.uk/news/uk-northern-ireland-32888240

[17] (2014). Royal Victoria Hospital: Delays ‘Contributed to Five Deaths. Accessed: Oct. 14, 2015. [Online]. Available: http:// www.bbc.co.uk/news/uk-northern-ireland-26135756

[18] J. P. Ruger, L. M. Lewis, and C. J. Richter, ‘‘Identifying high-risk patients fortriageandresourceallocationintheED,’’Amer.J.Emerg.Med.,vol.25, pp. 794–798, Sep. 2007, doi: 10.1016/j.ajem.2007.01.014.

[19] W.-T. Lin, S.-T. Wang, T.-C. Chiang, Y.-X. Shi, W.-Y. Chen, and H.-M. Chen, ‘‘Abnormal diagnosis of emergency department triage explored with data mining technology: An emergency department at a medicalcenterinTaiwantakenasanexample,’’ExpertSyst.Appl.,vol.37, no. 4, pp. 2733–2741, 2010, doi: 10.1016/j.eswa.2009.08.006.

[20] M. R. Baumann and T. D. Strout, ‘‘Triage of geriatric patients in the emergency department: Validity and survival with the emergency severity index,’’ Ann. Emerg. Med., vol. 49, no. 2, pp. 234–240, 2007, doi: 10.1016/j.annemergmed.2006.04.011.