Int. j. inf. tecnol. (June 2023) 15(5):2613–2621
https://doi.org/10.1007/s41870-023-01296-6

ORIGINAL RESEARCH

Adoption of machine learning algorithm for predicting the length of stay of patients (construction workers) during COVID pandemic

S. Selvakumara Samy[1] · S. Karthick[1] · Meghna Ghosal[1] · Sameer Singh[1] · J. S. Sudarsan[2] · S. Nithiyanantham[3]

Received: 16 November 2022 / Accepted: 15 May 2023 / Published online: 9 June 2023

Abstract The construction sector in a rapidly developing country like India is a very unorganized sector. A large number of workers were affected and hospitalized during the pandemic. This situation is costing the sector heavily in several respects. This research study was conducted as part of using machine learning algorithms to improve construction company health and safety policies. LOS (length of stay) is used to predict how long a patient will stay in a hospital. Predicting LOS is very useful not only for hospitals, but also for construction companies to measure resources and reduce costs. Predicting LOS has become an important step in most hospitals before admitting patients. In this post, we used the Medical Information Mart for Intensive Care(MIMIC III) dataset and applied four different machine learning algorithms: decision tree classifier, random forest, Artificial Neural Network (ANN), and logistic regression. First, I performed data pre-processing to clean up the dataset. In the next step, we performed function selection using the Select Best algorithm with an evaluation function of chi2 to perform hot coding. We then performed a split between training and testing and applied a machine learning algorithm. The metric used for comparison was accuracy. After implementing the algorithms, the accuracy was compared. Random forest was found to perform best at 89%. Afterwards, we performed hyperparameter tuning using a grid search algorithm on a random forest to obtain higher accuracy. The final accuracy is 90%. This kind of research can help improve health security policies by introducing modern computational techniques, and can also help optimize resources.

Keywords Length of stay · Hospital · Machine learning · Computer science · Random forest · Logistic regression · Decision tree · k-Nearest neighbors (KNN)

1 Introduction

LOS (length of stay) is the estimation given by the hospital to the patient regarding the time the patient is most likely to spend in the hospital given his/her condition, the duration changes with diagnosis, severity of condition, constitution of the patient and various other factors. Estimating the LOS at the time of admission offers anguesstimate [1, 2] of the patient's release date, which assists in the planning of critical medical operations. LOS is also considered to be a reliable proxy for measuring the consumption of hospital resources. LOS is important from the patients point of view because

✉ S. Nithiyanantham
s_nithu59@rediffmail.com

[1] Department of Computational Intelligence, SRM Institute of Science and Technology, Kattankulathur, Tamilnadu 603203, India

[2] School of Energy and Environment, NICMAR University, 25/1, Balewadi, Pune 411045, India

[3] Department of Physics, (Ultrasonic/NDT and Bio-Physics Divisions), Thiru. Vi. Kalyanasundaram Government Arts and Science College (Affiliated to Bharathidasan University, Thiruchirapalli), Thiruvarur, Tamilnadu 610003, India

it's also helpful in estimating the total cost of treatment, and gives the patients extra time to plan and facilitate finance [3].

To further explain the importance of LOS let's consider a scenario without it, patients will be told the final price at the time of discharge, a certain percentage of people will require time to come up with the money which will cause delays and losses, the percentage of people that dispute the bill in court will increase by a big margin both of which will increase the hospitals expenditure and as a result the hospital will charge more money from the patients to recover the losses and that in turn will cause more disputes and the cycle will continue[4–6].

The length of stay can also act as an significant indicator for judging the effectiveness of hospital management which can be used by the hospital to improve their management for swifter and efficient treatment, it can be a tool to test new polices and their effects [7–9]. LOS can be used by the patients to compare and contrast between different hospitals to choose from. All of this will result in increased efficiency and a reduction in the number of days of stay, which will reduce the risk of infection and drug side effects while also increasing treatment quality. As a result, predicting the LOS value at the time of patient admission is crucial for designing a hospital logistical planning strategy [10–13].

India is the second largest population in the world and the 7th largest in terms of area, just this disparity in these numbers puts a large burden and a huge significance on the construction industry and as far as economy is concerned, the ups and downs of the industry is an important indicator for the development of the county as it creates opportunity to create wealth, and also is a big source of revenue for the government in the form of taxes.

The construction industry has over 8 percent stake in the GDP, that puts it over 6 hundred thousand crores in value, and if we take into account the amount of job opportunities it creates and how many government policies and schemes use it as means to help and employ people, it becomes an invaluable and indispensable asset to the country.

In a developing country with ever expanding infrastructure like India construction industry is one of the most necessary industries of the country, the Indian construction industries majorly comprises of small enterprises, about 95 percent and majority of its worker fall under the unorganized sector that means they don't have any benefits, perks or security that is enjoyed by the people that are employed in the organized sector. Over 80 percent of the people who work as construction workers work on a "no work, no pay" regime, that means if they don't work for a single day they might not eat that day, and these people are one of the groups that has suffered the most because of the COVID 19 pandemic. The official maximum allowed work hours for workers is 8 but more often than not these people have to work more than 12 h just to afford food for their families.

COVID-19 is a global infectious disease outbreak that has infected and killed around 550,000 individuals. It's a coronavirus that may infect both animals and people. This one began at the end of 2019 in Wuhan, China, a metropolis of 11 million people. Coronavirus outbreaks have been a source of concern over the past 2 decades.

Some of the common COVID-19 symptoms are fever, cold, sore throat, difficulty in breathing, loss of taste and smell. The measures to be followed by every individuals to reduce the spread of COVID 19 included wearing a mask all the time, maintain social distancing, washing hands with a soap or a sanitizer.

The LOS is a complicated variable that is affected by a variety of circumstances, including the medical and community setting of the patient and his treatment [14, 15]. This type of service to which the patients are admitted influences the prognosis of LOS. This models used in an extremity division differ from those used in a scheduled department. Considerate the elements that influence the LOS is critical for predicting its value [16–18]. In this field of study, artificial intelligence approaches such as Machine Learning and data mining are applied. First, analyse the elements influencing LOS, Based on these characteristics, develop a LOS prediction model [19–21].

Our aim is to help the construction workers and the hospitals to predict LOS before admission. Whenever a patient is admitted, the hospitals will collect their medical history and medical problems [22–24]. Then they will check their database and search for patients with similar medical history or problems. Then will use the Random Forest algorithm and will use the information of the similar kinds of patients to predict the length of stay for the particular patient [25–27].

This is Classification Problem. So we have used the 4 classification algorithms [28]. Those are: Random Forest, Decision Tree, KNN and Logistic Regression [29–31] as shown in Fig. 1.

2 Data set description

For our study, we used the Medical Information Mart for Intensive Care (MIMIC III) data collection. It stands for

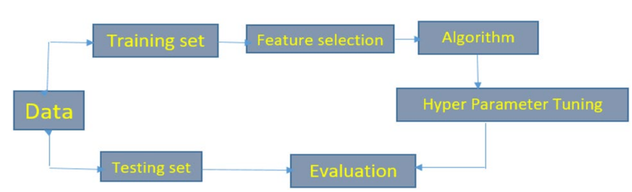

Fig. 1 Proposed model architecture

the dataset includes information collected from over large number of patients admitted in ICU. The dataset contains 58,976 rows and 18 columns. The MIMIC data set is an open source data set that contains 2 types of data: categorical data and numerical data. The data set contained some unwanted columns such as gender, marital status, religion, ethnicity etc. which were removed during data pre-processing. After performing encoding for the target column (length of stay) 0 represents 1–4 days, 1 represents 5–8 days, 2 represents 9–12 days and 3 represents 13–16 days.

There were multiple attributes that were taken under consideration for the study. Some of them are: Notes, lab tests, and liquidstability are kept in a sequence of 'events' characteristics.For example, the output events database contains all output measures for a certain patient, whereas the lab events table contains laboratory test results for a patient.chart events is associated with a single itemid that represents the concept being measured but does not include the measurement's name. admissions, patients, icustays, services, and transfers are the five tables used to define and track patient stays. The dictionaries D CPT, D ICD DIAGNOSES, D ICD PROCEDURES, D ITEMS, and D LABITEMS are used to compare codes to their meanings. The subsequent tables provide information on patient care such as physiological measurements, carer observations, and billing data.

3 Literature review

See Table 1.

4 Methodology

4.1 Data pre-processing

It is an significant step in order to make correct predictions with high accuracy rate. The data pre-processing includes cleaning the data set. At first the missing values were assigned and then the duplicated rows and unwanted columns such as gender, insurance, marital status, ethnicity were removed.

4.2 Encoding

When working with categorical data in machine learning, encoding is a needed pre-processing step.In ML all the input and output variables must in numeric format. The data set that was used contained mostly categorical data. We converted the categorical data to numeric format so that we can fit and evaluate our model.

4.3 Feature selection

Feature selection is required in order to reduce the number of input variables and choose the best features for achieving high accuracy. SelectKBest Algorithm selects the best features based on a particular score function and a k value. The score function used for our project is chi2 and the k value is 16 which were selected based on a trial error method. With k = 16 we got the highest accuracy as shown in Fig. 2.

Before feature extraction we had 28 featuresas shown in Figs. 3, 4. After that it was reduced to 16 as shown in Fig. 5 as shown in Table 1.

4.4 Building the model

For the training and testing part, the data set was divided in such a way that 80% of the data set was used for training and 20% of the data set was used for the purpose of testing.

4.4.1 ML algorithms used

We were used four Machine Learning Algorithms.

4.4.1.1 Decision tree classifier It is a machine learning algorithm that is supervised. A test relates to the internal node. Figure 6 depicts each branch according to the test result and each leaf node assigning a categorization. As an example: Consider the data set below, where H represents high, L represents low, and N represents normal. Table 2 shows A and B, which reflect the drugs/medicines consumed by the patient.

4.4.1.2 Logistic regression LR is a supervised machine learning approach that may be used to solve both classification and regression problems, though it is more commonly used for classification. It solves both binary and linear classification problems. It performs good and achieves high accuracy when working with linearly separable classes. There are 3 types of logistic Regression:

Binary logistic regression: It is used when there are only 2 possible outcome. For eg: yes/no, 0/1.

Multinomial logistic regression: It is used when there are more than 2 classes.

Ordinal logistic regression: It is used when there are more than 2 classes with ordering. For eg: ranking (1–10).

4.4.1.3 KNN KNN is an algorithm that can be used for both regression and classification based on supervised learning technique, in this method the new values classification is

Table 1 Final 16 feature extraction algorithm

S.No	Authour	Title	Source	Findings
1	Khosravizadeh, O., Vatankhah, S., Bastani, P., Kalhor, R., Alirezaei, S., Doosty, F	"Factors affecting length of stay in teaching hospitals of a middle-income country"	Electron. Phys. 8(10), 3042–3047 (2016)	Khosravizadeh et al. have shown that age, work, marital status, past hospitalisation history, patient condition at discharge, form of payment, and kind of therapy may all affect LOS. Because health-care resources are limited, especially in hospitals, identifying these features can help reduce needless LOS and maximise resource usage
2	Carter, E.M., Potts, H.W.:	Predicting length of stay from an electronic patient record system: "a primary total knee replacement example	BMC Med. Inform. Decis. Mak. 14(1), 1–13 (2014)	Carter, E.M., and colleagues proved that their technique can be used for resource planning and particular patient expectations of their LoS for PTK patients admitted to any hospital in any country. The model, as well as the decision support tool (created in Excel), are straightforward to implement within this hospital's data warehouse
3	Turgeman, L., May, J.H., Sciulli, R.:	Insights from a machine learning model for predicting the hospital length of stay (LOS) at the time of admission	Expert System Applications 78, 376–385 (2017)	Turgeman, L. et al. employed a regression tree (Cubist) model to predict LOS using static inputs, or data that is known at the time of admission and does not change during the patient's hospital stay. The model was trained and validated using de-identified administrative data from Veterans Health Administration (VHA) institutions in Pittsburgh, PA. They selected a cubic model because it predicted more accurately than other techniques. Furthermore, tree models enable us to analyse the data-derived classification criteria in order to better identify the components most linked to hospital LOS
4	Shea, S., Sideli, R.V., Dumouchel, W., Pulver, G., Arons, R.R., Clayton,	"Computer-generated informational messages directed to physicians": effect on length of hospital stay"	P.D. J. Am. Med. Inform. Assoc. 2(1), 58–64 (1995)	According to Shea, S., Sideli, et al., proposed hospital payment by diagnosis-related group (DRG) and length of stay (LOS) has become a critical problem in hospital attempts to minimise expenditures. Because Columbia-Presbyterian Medical Center (CPMC) has above-average LOSs for several DRGs, the authors hypothesised that a computer-generated instructional message directed at clinicians would decrease LOS
5	Rowan, M., Ryan, T., Hegarty, F.,	"The use of artificial neural networks to stratify the length of stay of cardiac patients based on preoperative and initial postoperative factors"	Hare, N.OArtificial Intelligence Medicine. 40, 211–221 (2007)	ANNs, especially ensembles of ANNs, are effective for outcome prediction tasks in postoperative cardiac patients

Table 1 (continued)

S.No	Authour	Title	Source	Findings
6	Lafaro, R.J., Pothula, S., Kubal, K.P., Inchiosa, M.A., Pothula, V.M., Yuan, S.C., Maerz, D.A., Montes, L., Oleszkiewicz, S.M., Yusupov,	"Neural network prediction of ICU length of stay following cardiac surgery based on pre- incision variables"	A., Perline, R., Inchiosa, M.A Plos One 10(12), 1–19 (2015)	According to Rowan, M. et al. R.J. Lafaro et al. compared a neural network model to several other models in predicting the length of stay (LOS) in the heart surgery intensive care unit (ICU) based on pre-incision patient data in their study. Methods Thirty-six parameters were evaluated in 185 heart surgery patients to see how they affected ICU LOS
7	Chuang, M.T., Hu, Y.H., Tsai, C.F., Lo, C.L., Lin, W.CIn:	"The identification of prolonged length of stay for surgery patients"	Proceedings–2015 IEEE International Conferenceon Systems, Man, and Cybernetics, SMC 2015, pp. 3000–3003 (2016)	Chuang, M.T. et al. examined the whole history, medical records, and lab data of 897 clinical patients for whom general surgery practitioners performed operations. These clinical cases were divided into two groups: urgent procedures (462 instances) and non-urgent surgeries (434 instances) in order to build a prolonged LOS prediction model using various supervised learning algorithms. The random forest approach provided the most accurate and stable prediction model, according to the results

predicted based on similar available data and for regression average is taken from the similar available data.

KNN doesn't make any assumption using the available data hence it is a non-parametric algorithm, instead of actually learning from the data set it performs various computations on the available data set and makes the prediction, it's called a lazy learner algorithm. When a new data point is introduced it simply classifies it into a category which is common to data points which are similar to it, at the time of learning it simply stores the data.

4.4.1.4 Random forest RF is a classification and regression technique that employs Ensemble Learning and is primarily based on the notion of decision tree algorithms. Multiple Decision Trees are employed, and the outcomes are tallied.

Two methods are used in Random Forest. One is Bagging and another is Boosting.

Bagging: Bootstrapping the data and using it's aggregate to make a decision is known as Bagging. In this case a bunch of individual models are trained parallel and each model is trained by a subset of data.

Boosting: In this case the models are trained sequentially that is one after the other. New models learn from its previous models.

4.5 Hyperparameter tuning

After comparing the accuracies of the models it was noticed that Random Forest performed the best. Then we performed hyper parameter tuning using Grid Search CV to get the best parameters for our modelare shown in Figs. 6, 7. The parameters that we got are:

5 Data visualization

5.1 Number of patients admitted for a given range of time

From Figure 8. Below is representing the total number of patients in each time interval. 0 represents 1–4 days, 1 represents 5–8 days, 2 represents 9–12 days and 3 represents 13–16 days. From the graph we can observe that majority of the people were admitted for 5–8 days.

5.2 Number of patients admitted under a particular category from Fig. 9

Patients were admitted under the emergency category.

Age	NumCallouts	NumDiagnosis	NumProcs	NumCPTevents	Numinputs	NumLabs	NumMicroLabs	NumNotes	NumOutput	NumRx	NumProcEvents	NumTransferd	NumChartEvents	TotalNum	NewBorn
35	0.16	2.59	0.00	1.30	25.12	43.44	0.65	0.05	5.19	14.91	1.13	0.65	398.7	493.89	0
59	0.25	2.23	0.99	1.98	13.61	43.44	1.24	1.59	5.45	7.18	0.99	1.24	373.02	465.71	0
48	0.0	0.75	0.17	0.83	11.49	55.94	0.33	0.15	4.14	6.23	0.0	0.33	286.21	344.00	0
73	0.41	0.69	0.27	0.69	20.30	33.39	0.69	0.17	9.05	11.52	0.0	0.96	526.06	603.05	0
60	0.0	3.69	0.82	2.25	20.49	32.24	0.61	0.34	16.19	25.00	2.85	2.05	554.92	679.84	0

```
s = SelectKBest(score_func=chi2, k=16)
best= s.fit_transform(X,y)
```

Fig. 2 Feature selection using SelectKBest algorithm

```
print(X.columns)

Index(['age', 'admit_type', 'admit_location', 'NumCallouts', 'NumDiagnosis',
       'NumProcs', 'NumCPTevents', 'NumInput', 'NumLabs', 'NumMicroLabs',
       'NumNotes', 'NumOutput', 'NumRx', 'NumProcEvents', 'NumTransfers',
       'NumChartEvents', 'TotalNumInteract'],
      dtype='object')

# One hot encoding

category = [
                    'gender',
                    'admit_type',
                    'admit_location'
                      ]

for c in category:
    if c in X.columns:
        oneh = pd.get_dummies(X[c])
        X = X.drop(c, axis=1)
        X = X.join(oneh, lsuffix='_left', rsuffix='_right')

X.head()
print(X.shape)

(58976, 28)

from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2, f_regression
```

Fig. 3 Feature extraction before using SelectKBest algorithm

```
# Feature selection

s = SelectKBest(score_func=chi2, k=16)
best= s.fit_transform(X,y)

print(best.shape)

(58976, 16)
```

Fig. 4 Feature extraction after using SelectKBest algorithm

5.3 Number of patients in each admit location

From Fig. 10. We can observe the number of patients in each location.

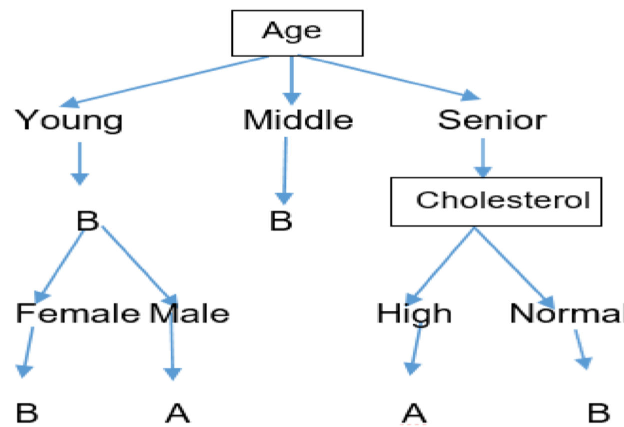

Fig. 5 Decision tree

```
scoring='accuracy'
model = RandomForestClassifier()

para = [{},]

grid_search = GridSearchCV(model, para, cv=5, scoring=scoring)
grid_search.fit(X, y)

print(grid_search.best_estimator_)
```

Fig. 6 Accuracy using random forest

Table 2 Decision tree

Id	Age	Gender	BP	cholesterol	Drug
1	Young	Female	H	Normal	A
2	Middle	Female	L	High	B
3	Senior	Male	N	High	A
4	Young	Male	L	Normal	B

```
RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
                       max_depth=None, max_features='auto', max_leaf_nodes=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, n_estimators=10,
                       n_jobs=None, oob_score=False, random_state=None,
                       verbose=0, warm_start=False)
```

Fig. 7 Hyper parameter tuning

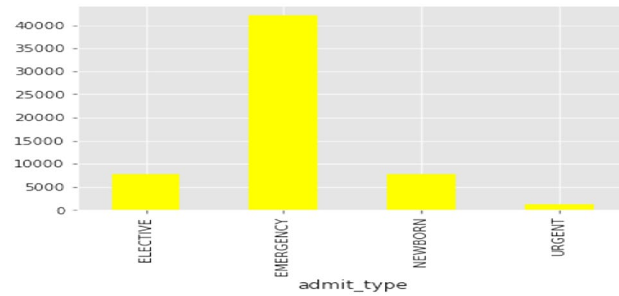

Fig. 9 Patients were admitted as emergency cases

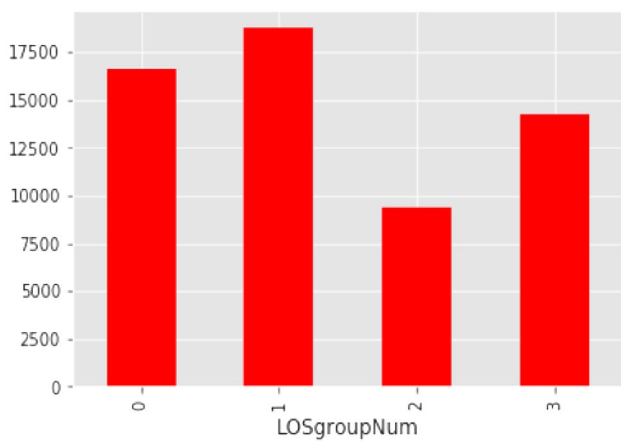

Fig. 8 Total number of patients in each time interval

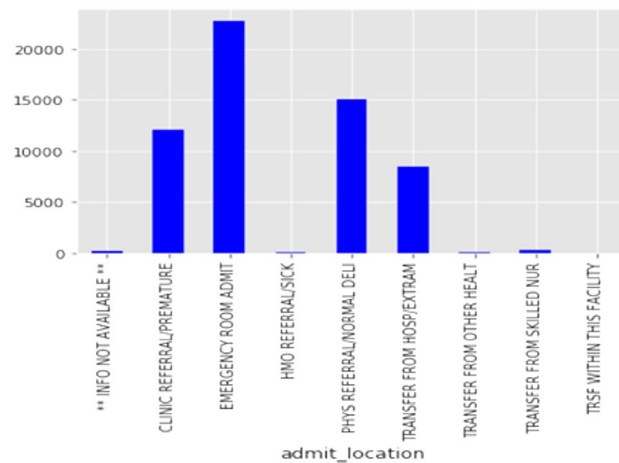

Fig. 10 Number of patients in each admit location

6 Results

After performing the data pre-processing and feature selection we implemented 4 Machine Learning algorithms and compared their accuracy (Fig. 11). The results that we got are as shown in Fig. 12 and Table 3.

We achieved the highest accuracy by Random Forest. Then we performed hyper parameter tuning to improve the accuracy.

6.1 Random forest

We used train and test dataset to get 89 percentage as shown in Fig. 12.

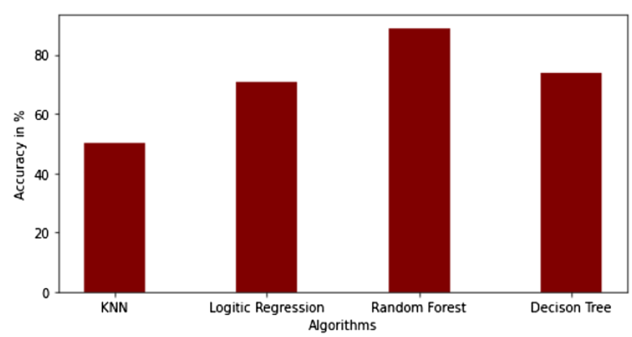

Fig. 11 Machine learning algorithms and compared their accuracy

6.2 Decision tree

We used train and test dataset to get 74 percentage as shown in Fig. 13.

```
Train: 0.9951886392539212
Test: 0.8961512377076976
```

Fig. 12 Train and test dataset for random forest

Table 3 Accuracy

Algorithms	Accuracy (%)
Logisic regression	71
Random forest	**89**
Decision tree	74
KNN	50

Random forest algorithm showing maximum accuracy is highlighted in bold

```
Train: 0.7450614667231877
Test: 0.7420311970159376
```

Fig. 13 Train and test dataset for decision tree

```
Train: 0.7253497244595167
Test: 0.7184638860630722
```

Fig. 14 Train and test dataset for logistic regression

6.3 Logistic regression

We used train and test dataset to get 71percentage as shown in Fig. 14.

6.4 KNN

We used train and test dataset to get 50 percentage as shown in Fig. 15.

7 Conclusion

The aim of this study was to predict a patient's duration of stay in a hospital. We analyzed the data set to select the best features for our model. We considered how to improve the working and management of the hospitals by predicting LOS. LOS is very important for the hospitals to manage it's resources and patients. It is also necessary to improve the financial structure by managing their requirements and to reduce the medical costs of the patient by reducing their length of stay.

```
Train: 0.6146672318779144
Test: 0.5053407934893184
```

Fig. 15 Train and Test dataset for KNN

8 Future enhancement

The built model and gathered data will be used in the future to estimate and anticipate the requirement for nasal oxygen assistance for COVID-19 patients in order to better manage these resources. Furthermore, the constructed model may be utilized with a bigger dataset derived from the core MIMIC III data set to generalize the model to the kingdom or globally.

Data Availability The data can be made available proper request from the editor.

References

1. Khosravizadeh O, Vatankhah S, Bastani P, Kalhor R, Alirezaei S, Doosty F (2016) Factors affecting length of stay in teaching hospitals of a middle-income country. Electron Phys 8(10):3042–3047
2. Carter EM, Potts HW (2014) Predicting length of stay from an electronic patient record system: a primary total knee replacement example. BMC Med Inform Decis Mak 14(1):1–13
3. Turgeman L, May JH, Sciulli R (2017) Insights from a machine learning model for predicting the hospital Length of Stay (LOS) at the time of admission. Expert System Appl 78:376–385
4. Shea S, Sideli RV, Dumouchel W, Pulver G, Arons RR, Clayton PD (1995) Computer-generated informational messages directed to physicians: effect on length of hospital stay. J Am Med Inform Assoc 2(1):58–64
5. Rowan M, Ryan T, Hegarty F, Hare NO (2007) The use of artificial neural networks to stratify the length of stay of cardiac patients based on preoperative and initial postoperative factors. Artif Intell Med 40:211–221
6. Lafaro RJ, Pothula S, Kubal KP, Inchiosa MA, Pothula VM, Yuan SC, Maerz DA, Montes L, Oleszkiewicz SM, Yusupov A, Perline R, Inchiosa MA (2015) Neural network prediction of ICU length of stay following cardiac surgery based on preincision variables. PLoS ONE 10(12):1–19
7. Chuang MT, Hu YH, Tsai CF, Lo CL, Lin WC (2016) The identification of prolonged length of stay for surgery patients. In: Proceedings 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015, pp 3000–3003
8. Loshin D (2009) Data consolidation and integration. Master data management. Elsevier, pp 177–199
9. Li JS, Tian Y, Liu YF, Shu T, Liang MH (2013) Applying a BP neural network model to predict the length of hospital stay. LNCS, 7798th edn. Springer, pp 18–29
10. Pendharkar PC, Khurana H (2014) Machine learning techniques for predicting hospital length of stay in Pennsylvania federal and specialty hospitals. Int J Comput Sci Appl 11:45–56
11. Johnson AE, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, Moody B, Szolovits P, Anthony Celi L, Mark RG (2016)

MIMIC-III, a freely accessible critical care database. Sci Data 3:1–9
12. Microsoft (2017) Predicting hospital length of stay
13. Aghajani S, Kargari M (2016) Determining factors influencing length of stay and predicting length of stay using data mining in the general surgery department. Hosp Pract Res (HPR) 1(2):53–58
14. Vijayan K, Ramprabu G, Selvakumara Samy S, Rajeswari M (2020) Cascading model in underwater wireless sensors using routing policy for state transitions. Microprocess Microsyst 79:1098
15. Arun C, Karthick S, Krishnan SM, Dutta GS (2019) A tool for analyzing software requirements document quality. Int J Recent Technol Eng 8(2 Special Issue 4):274–279
16. Ramraj S, Karthick S, Yashwant K (2018) Application of machine learning for renewable energy prediction. Indian J Public Health Res Dev 9(9):1048–1053
17. Adjuik TA, Ananey-Obiri D (2022) Word2vec neural model-based technique to generate protein vectors for combating COVID-19: a machine learning approach. Int J Inf Technol 14:3291–3299
18. Karthick S, Malathi D, Sudarsan JS, Nithiyanantham S (2021) Performance, evaluation and prediction of weather and cyclone categorization using various algorithms. Model Earth Syst Environ 7(3):1703–1711
19. Selvakumara Samy S, Sivakumar V, Sood T, Negi YS (2023) Intelligent web-history based on a hybrid clustering algorithm for future-internet systems. Artificial intelligence and evolutionary computations in engineering systems. Springer
20. Sureshkumar S, Mathan G, Ri P et al (2022) Deep learning framework for component identification. Int J Inf Technol 14:3301–3309
21. Karthick S, Arun C, James SJ, Selvakumarasamy S (2023) Virtual personal/voice based assistant for an institution through Alexa's skill. In: Materials Today: Proceedings
22. James HVJ, Selvakumara Samy S (2023) Advanced encryption using fuzzy logic and secret sharing scheme. Int J Pure Appl Math 118(22):1743–1748
23. Aziz RM, Baluch MF, Patel S et al (2022) LGBM: a machine learning approach for Ethereum fraud detection. Int J Inf Technol 14:3321–3331
24. Latha P, Selvakumara Samy S (2021) Emotion detection in IoT-based E-learning using convolution neural network. Fuzzy intelligent systems: methodologies, techniques, and applications. Wiley, pp 27–44
25. Selvakumara Samy S, Latha P (2018) A novel feature extraction approach using principal component analysis and quantum behaved particle swarm optimization-support vector networks for enhancing face recognition. J Comput Theor Nanosci 15(9–10):3012–3016
26. Selvakumara Samy S, Latha P (2016) A novel approach for E-learning using QPSO algorithm. Int J Control Theory Appl 9(28):79–86
27. Aksharan MK, Priya B, Selvakumara Samy S, Haribaabu V (2019) Emotion recogination using LDA and CNN in E-learning. Int J Adv Sci Technol 28(7):370–378
28. Selvakumara-Samy S, Vikash K, Rishabh RS (2018) Augmented reality in E-learning system. Int J Pure Appl Math 118:291–296
29. James SJ, Haribaabu M, Selvakumara S (2016) Anaglyph 3D image encryption using multiple meaningful secret hiding scheme and modified error diffusion. Int J Control Theory Appl 9(28):261–275
30. Sudarshan Anupama CG, Samy SK, Surya E (2016) Empirical analysis of software testing tools using evaluation criteria. Digit Libr 9(28):251–260
31. Sudarsan JS, Saravana Kumar R, Prasanna K, Karthick S, Suprajha GS (2020) Estimation of air pollution concentration in suburban of chennai city and validation of the same using air quality models. Artificial intelligence and evolutionary computations in engineering systems. Springer, pp 603–610