International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5571
***Article Received**: 05 May 2022 **Revised**: 30 May 2022 **Accepted**: 20 June 2022 **Publication**: 31 July 2022*

Predicting The Discharge of Patients Via Machine Learning Based Discharge Predictive Model

[1]Salma Fauzia, [2]Romana Anjum
[1,2] Muffakham Jah College of Engineering and Technology, Osmania University, Hyderabad, INDIA.

Abstract

The primary objective of the work is to create a discharge roster for patients by employing various machine learning techniques and to predict the discharge of a patient. The performance of the proposed discharge predictive model is measured through various performance measures. The research work is carried out based on the dataset formed with actual data of patients in hospital. The machine learning (ML) based Discharge Predictive Model is developed by combining well known ML algorithms like K-Nearest Neighbour (KNN) algorithm, Random Forests algorithm and Light Gradient Boosting algorithm with optimum parameters, various feature combinations and pre-processing techniques. The performance of the proposed model is measured in terms of accuracy and it is compared with various existing techniques like SVM and NN. The result of the comparison study exhibit that the proposed predictive learning model attained enhanced accuracy than other ML techniques.

Keywords: Machine learning, LGBM, Random Forests, and KNN.

1. Introduction

In the modern computerized world, almost every sector has become automated. The healthcare sector and disease diagnosis not an exception to this automation. The diagnosis of diseases becomes a lot quicker, easier and cheaper because of the usage of computer based diagnostic technologies. The time taken and efforts put involved in the decision-making process are also reduced. The automated computer diagnosis method is less vulnerable to the variability existing in the doctors' choices. Moreover, the computers are also helpful in creating the discharge summaries and discharge of patients, irrespective of the fact it is useful in the disease prediction. An improved model is employed by the most of the clinical decision-making tools. But one should verify whether the model is providing or resulting the better accuracy in decision making or not.

2. Statement

Machine learning is rapidly growing in the healthcare sector nowadays. The detection of factors like the nature of disease, treatment for a disease, length of stay in hospital and discharge of patients after the treatment requires an effective machine learning based prediction technique. The related works discuss about various machine learning techniques utilized in the healthcare sector. Most of the techniques makes use of certain flow and logistic regression for developing the prediction model. But the results of those techniques are not accurate and also demands a larger memory. Moreover, existing techniques comprises of various other limitations like complex in design, ineffective, requirement of skilled persons. Hence in order to rectify the limitations prevailing in the existing works and to predict the anaemic status of patient and discharge of a patient, a Machine Learning based Discharge Predictive Model is proposed in the research study.

3. Objectives

The objectives of the research work are listed below:

- To create and construct discharge roster for patients by employing various machine learning techniques.
- To predict the discharge of a patient admitted in a hospital.

4. Related Works

A multimodal disease risk prediction algorithm was proposed by (Cuadrado et al., 2019) based on the Convolutional Neural Network (CNN). On comparing with various conventional and existing prediction algorithms, the obtained accuracy of the proposed prediction algorithm was found to be 94.8% along with a convergence speed and found to be very much faster than the traditional unimodal disk prediction algorithm based on CNN.

(Jabbar, Samreen, Aluvalu, & Technology, 2018) discusses about the bright future of machine learning in the health care sector. How Machine Learning (MK) and

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5571
***Article Received**: 05 May 2022 **Revised**: 30 May 2022 **Accepted**: 20 June 2022 **Publication**: 31 July 2022*

Artificial Intelligence (AI) might change the healthcare in future was explained and about the necessity of rectifying the issues faced by doctors and patients in the diagnosis though ML and AI based decision support systems.

(Ferdous, Debnath, & Chakraborty, 2020) provided an extensive literature survey regarding various machine learning algorithms utilized in the sector of healthcare. A list of accuracy of various machine learning algorithms were given to indicate the prediction level of each algorithms.

(He et al., 2019) conducted a study based on which it was found that the simulation modelling is the major method to be utilized for studying about the inpatient bed management. Numerous opportunities were found to be available in various things like consideration of system-wide inpatient flow, employment of experimental techniques, integration of predictive techniques in the inpatient bed management and the optimization of inpatient bed allocation.

A mean error of lesser than half a day was attained by (Cuadrado et al., 2019) along with the coefficient of determination (R2) value being more than 97% for their prediction regarding ICU survivors or not survivors.

Though many existing works like (Pallathadka et al., 2021); (Mustafa & Rahimi Azghadi, 2021); (Jain & Chatterjee, 2020); (Chen et al., 2021); (Kaur, Sharma, & Mittal, 2018); (Vellido& applications, 2020); and (Siddique & Chow, 2021) aim to contribute to health care, still there exist the scope for serving it better by helping both the patients and health care professionals by letting know the discharge period.

5. Materials and Methods

We have proposed a health sector serving application-based project for overcoming the challenges suffered by the other conventional machine learning methodologies. The main contribution of our novel project is detecting and classifying various data to predict the discharge period with the help of the following machine learning classifiers:

- Random Forest
- KNN
- LGBM

For the implementation of the system, we have designed this project on a Python-oriented surrounding by using the flask structure.

The following are the aimed advantages out of our application-based project:

- Good Accuracy should be achieved.
- The project should be highly efficient enough to classify the discharge period.
- The complex level of the project should be lower.
- There should be no requirement of deploying proficient people.

The below fig. 1 is showing the block diagram of our application-based project aiming for classifying the discharge period.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5571
Article Received: 05 May 2022 Revised: 30 May 2022 Accepted: 20 June 2022 Publication: 31 July 2022

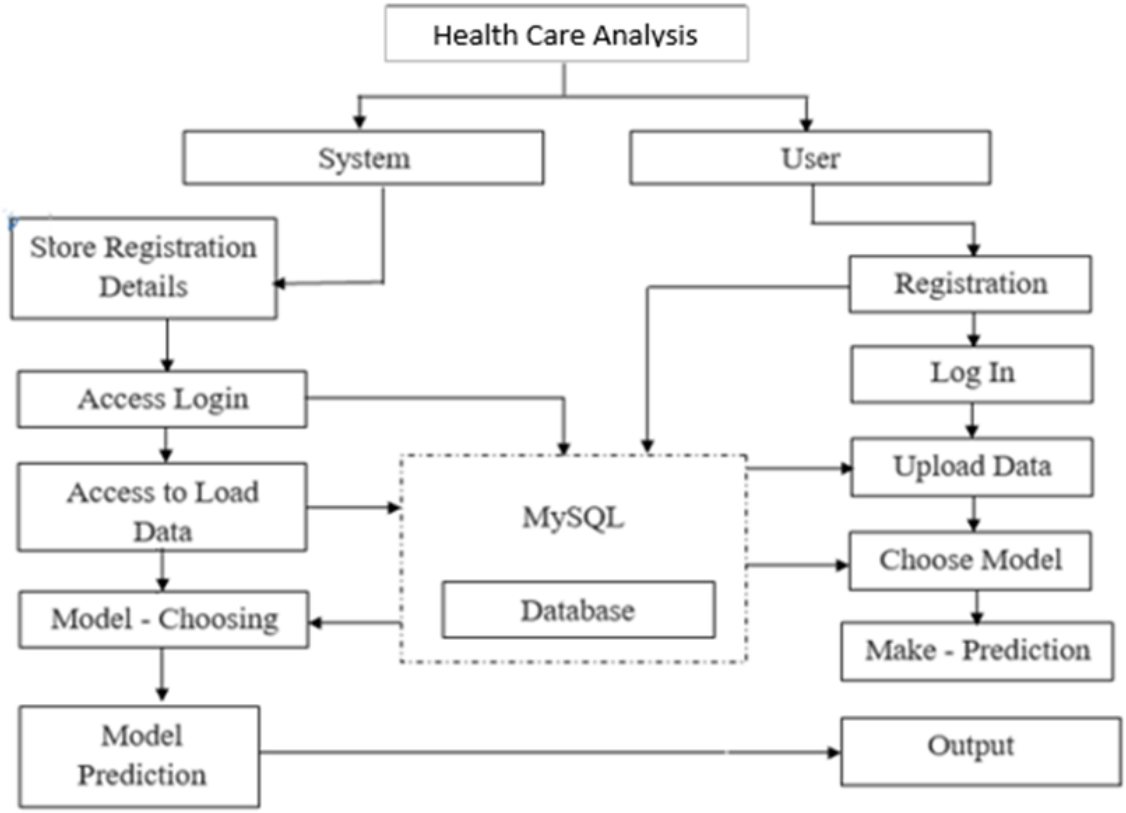


Fig. 1 Block Diagram of the Discharge Period classifier

Rather than using one decision tree, we have utilized the random forest classifier that contains numerous decision trees along with two other classifiers. For example, a training data consists of various observations of a phone and the features of the phone get split into 4 root nodes.

These root nodes describe 4 features like camera, price, RAM, and internal storage, which impact the choice of the user. The classifier will divide the nodes after the random selection of features. The end forecasting will be made depending upon on the output of those four trees.

The output selected by the majority of the decision among 4 decision trees will constitute the end choice. Purchase will be the choice of the customer in case the three-decision tree selects the purchase and remaining one selects don't purchase.

Using the KNN, we select k=5 (mostly preferred) and estimate the Euclidean distance intermediary to the data points. Thus, we know the nearest neighbours and there by decide further based on it.

LightGBM perform efficiently and contribute favourable storage utilization because of implementing a high optimal histogram-oriented decision tree method.

5.1 Review of our Software Development Life Cycle

In our project, we have made use of the waterfall model as our software development pattern. The model is shown in the below fig. 2.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5571
Article Received: 05 May 2022 Revised: 30 May 2022 Accepted: 20 June 2022 Publication: 31 July 2022

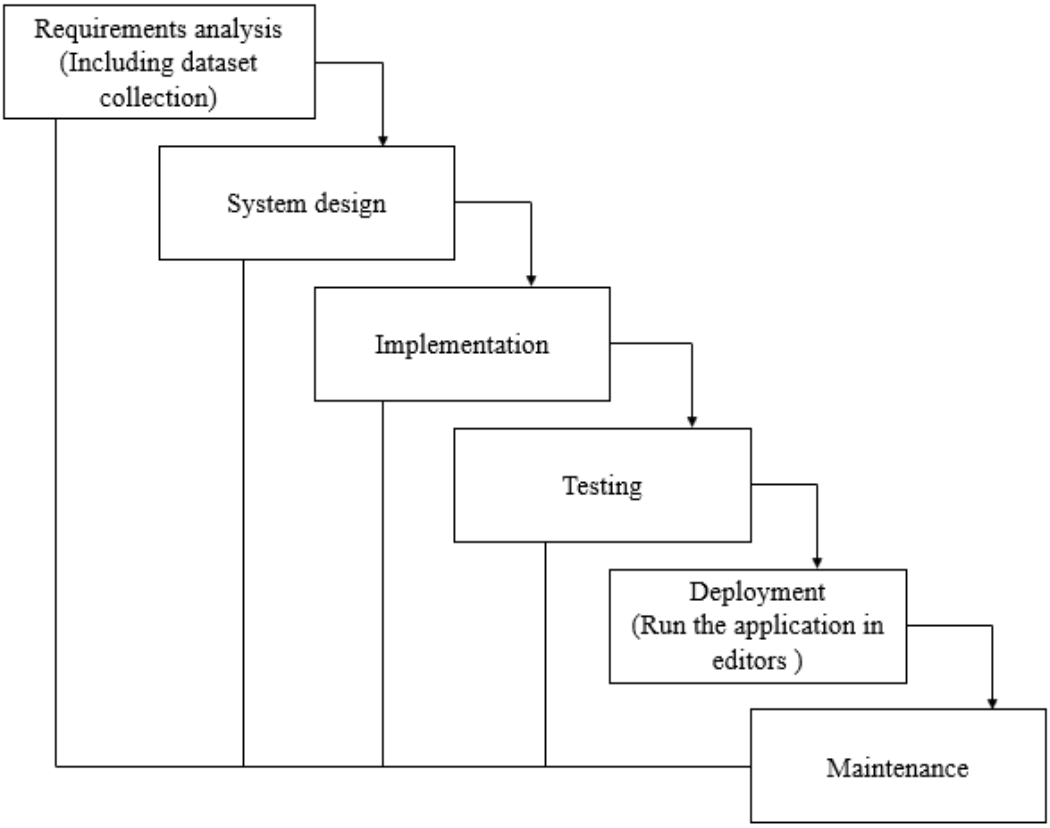


Fig. 2 Waterfall Model

5.2 Modules in our Project Design

The modules in our project are divided into two as follows:

5.2.1 Modules in the System

The following are the modules in the design of our system.

- **Consider dataset-** The system collects the data provided by the consumer.
- **Divide-** Divides the considered dataset into a ratio that is proportionate to the input provided by the consumer.
- **Training of Model-** The consumer then chooses the model depending upon on the parameters for getting a reply from the data and then to predict the outcomes from that model.
- **Produce outcomes-** The system could be able to convey the forecasted outcomes, which is shown to the consumer.

5.2.2 Modules for the Consumer

The following are the modules in the design for our consumer.

- **Enrolment-** Each newer consumer must enrol themselves in the system by using a distinct Mail ID and name.
- **Login-** Enrolled consumers then could get login in the system.
- **Transmit and View Data-** Once the consumer logs in successfully, the consumer could be able to transmit and view that transmitted dataset.
- **Divides and Choose Model-** The consumer can choose the preferred divide ratio in order to select the needed model for seeing the outcome yielded from the dataset.
- **Forecasting-** The consumer could forecast the output from the designed system.
- **Logoff-** Consumer gets logoff from the system.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5571
Article Received: 05 May 2022 Revised: 30 May 2022 Accepted: 20 June 2022 Publication: 31 July 2022

6. Performance Examination and Discussion

For examining the performance of the project, we use the following specifications:

6.1 System Specifications

The below are the software and hardware specifications that we have utilized for our project.

6.1.1 Software Specifications

Operating System	Windows 10
Server-side Script	Python 3.6
IDE	PyCharm
Libraries utilized	Pandas, NumPy, Scikit-Learn
Structure	Flask
Data Base	MySql

6.1.2 Hardware Requirements

Processor	i3- typed Intel Processor
RAM	8 GB (minimum)
Hard Disk	128 GB
Key Board	Standard Windows Keyboard
Mouse	Two or Three Button Mouse
Monitor	Any

6.2 Output Screenshots and Discussion

6.2.1 Home

In our project, we are identifying the anaemia on the basis of the details provided by the consumer. The home page shows the overview of the project to predict the discharge period in the health care sector.

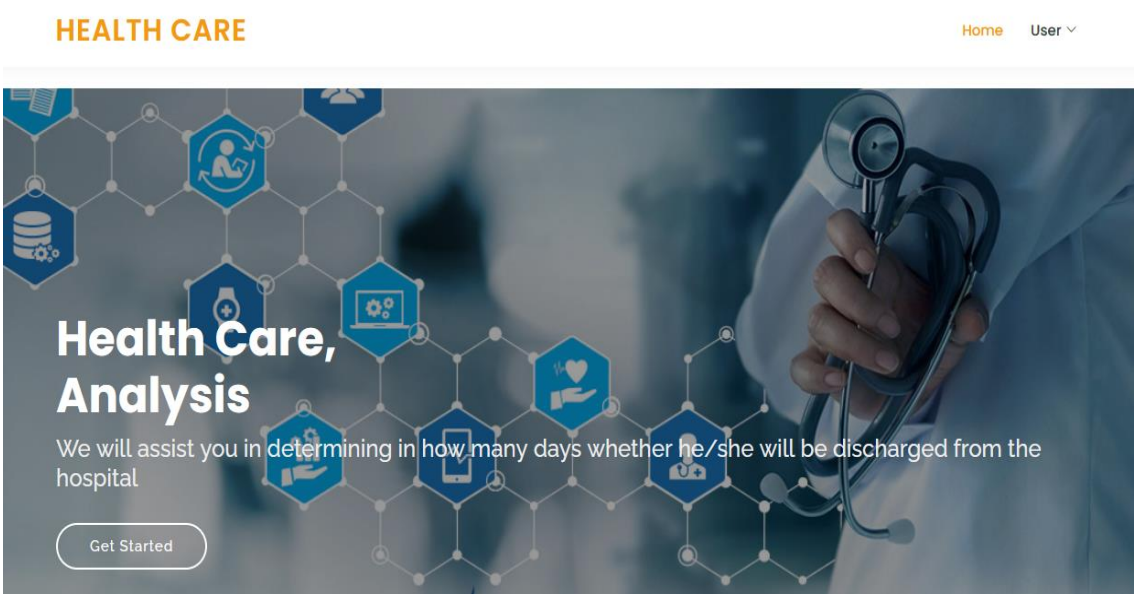


Fig. 3 Home Page of our project

6.2.2 About

In this page, the about page shows us the major objective for which this project was executed.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5571
Article Received: 05 May 2022 Revised: 30 May 2022 Accepted: 20 June 2022 Publication: 31 July 2022

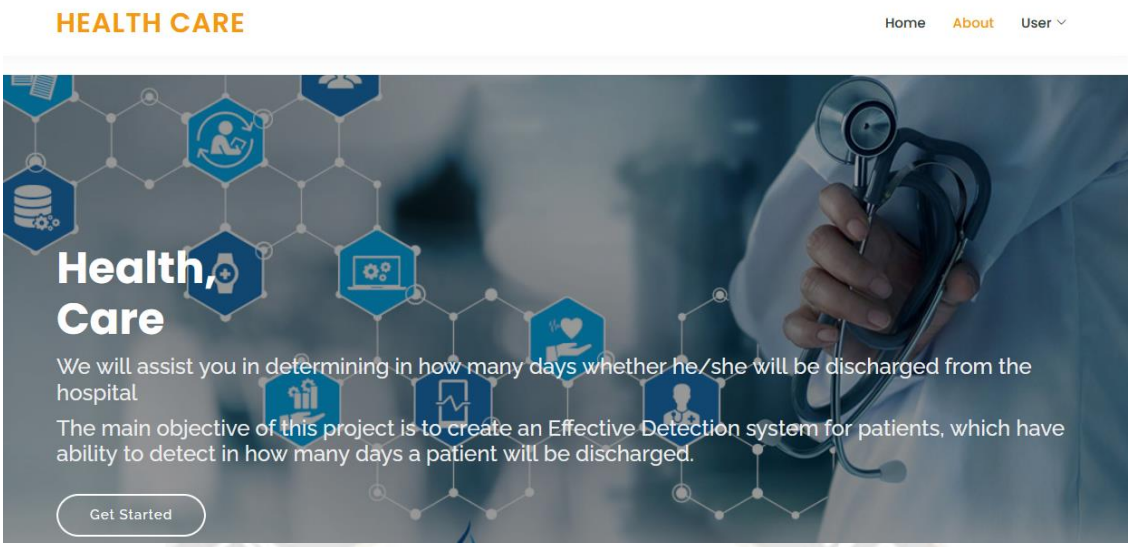


Fig. 4 About Page of our project

6.2.3 Registration

Registration page shows the fields for the consumer to fill details and get enrolled themselves for beginning.

Fig. 5 Registration Page of our project

6.2.4 Login

The login page prompts the consumer to enter his/ her enrolment details for getting logged into the system.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5571
Article Received: 05 May 2022 Revised: 30 May 2022 Accepted: 20 June 2022 Publication: 31 July 2022

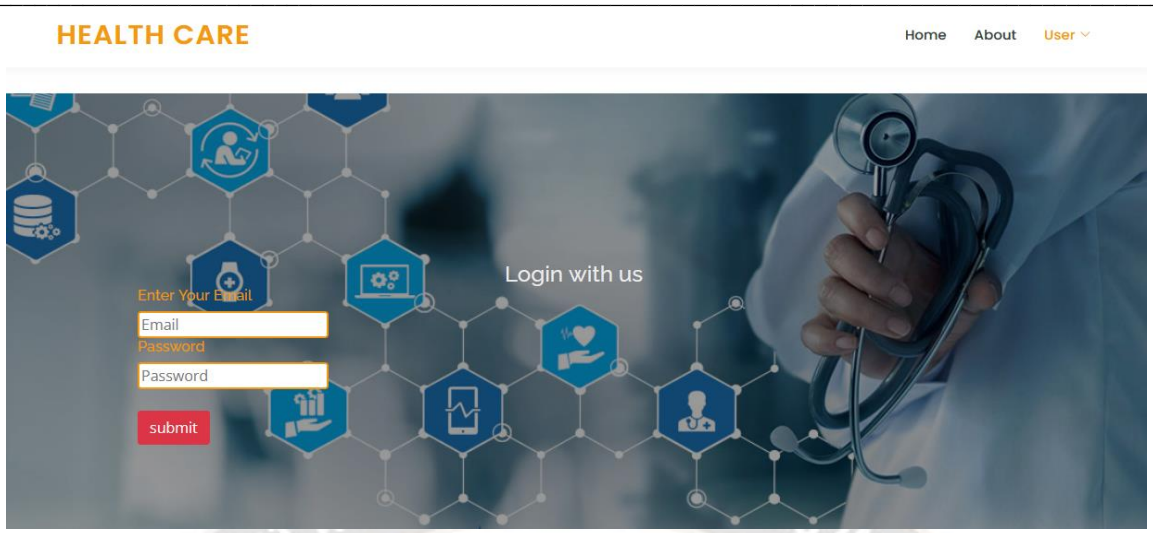


Fig. 6 Login Page of our project

6.2.5 Upload

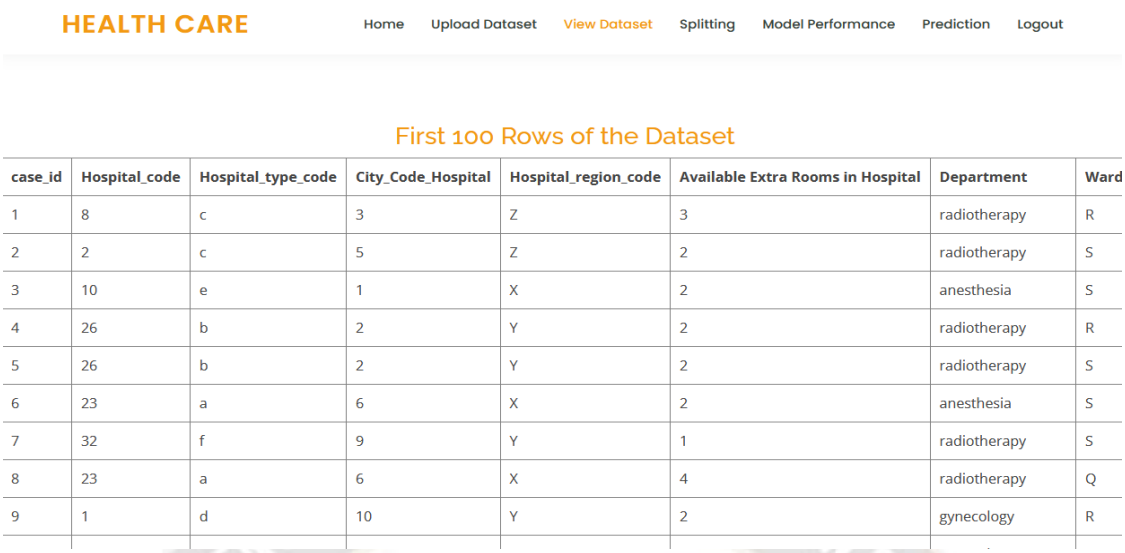
The upload page is empowered the consumer for uploading the needed dataset.

Fig. 7 Upload Page of our project

6.2.6 View Data

The consumer sees the data that he/ she had imported into the system.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5571
Article Received: 05 May 2022 Revised: 30 May 2022 Accepted: 20 June 2022 Publication: 31 July 2022



case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Available Extra Rooms in Hospital	Department	Ward
1	8	c	3	Z	3	radiotherapy	R
2	2	c	5	Z	2	radiotherapy	S
3	10	e	1	X	2	anesthesia	S
4	26	b	2	Y	2	radiotherapy	R
5	26	b	2	Y	2	radiotherapy	S
6	23	a	6	X	2	anesthesia	S
7	32	f	9	Y	1	radiotherapy	S
8	23	a	6	X	4	radiotherapy	Q
9	1	d	10	Y	2	gynecology	R

Fig. 8 View Dataset Page of our project

6.2.7 Split

This page provides the provision of splitting the data into train and test segments prior to the training of the model by entering the splitting size.

Fig. 9 Split Page of our project

6.2.8 Model Training

The model training gets completed here and then shows the score of the models.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5571
Article Received: 05 May 2022 Revised: 30 May 2022 Accepted: 20 June 2022 Publication: 31 July 2022

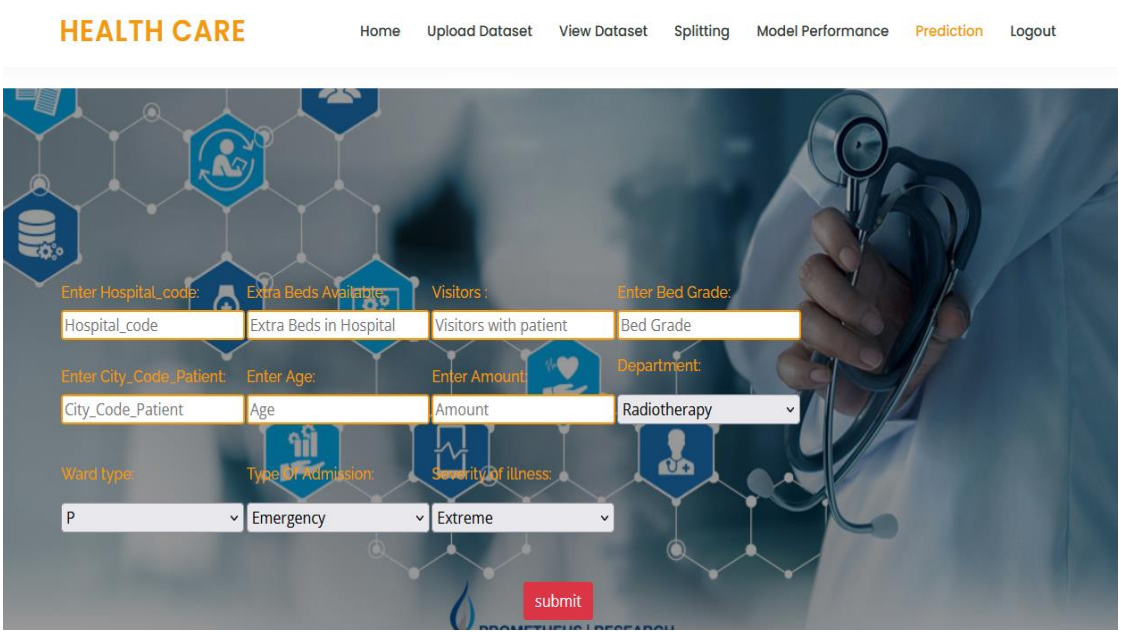
Fig. 10 Model Training Page of our project

Table 1. Model Performance of our project

Model Performance	Accuracy
Random Forest Algorithm	75.53%
K-Nearest Neighbour (KNN) Algorithm	47.05%
Light Gradient Boosting Machine (LGBM)	79.99%

6.2.9 Prediction

The consumer needs to provide the needed details to the various fields shown below for getting a valid reply from the data imported. The page will decide and say whether the patient is currently anaemic or non-anaemic for predicting the discharge period after getting hospitalized.



International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5571
Article Received: 05 May 2022 Revised: 30 May 2022 Accepted: 20 June 2022 Publication: 31 July 2022

Fig. 11 Prediction Page of our project

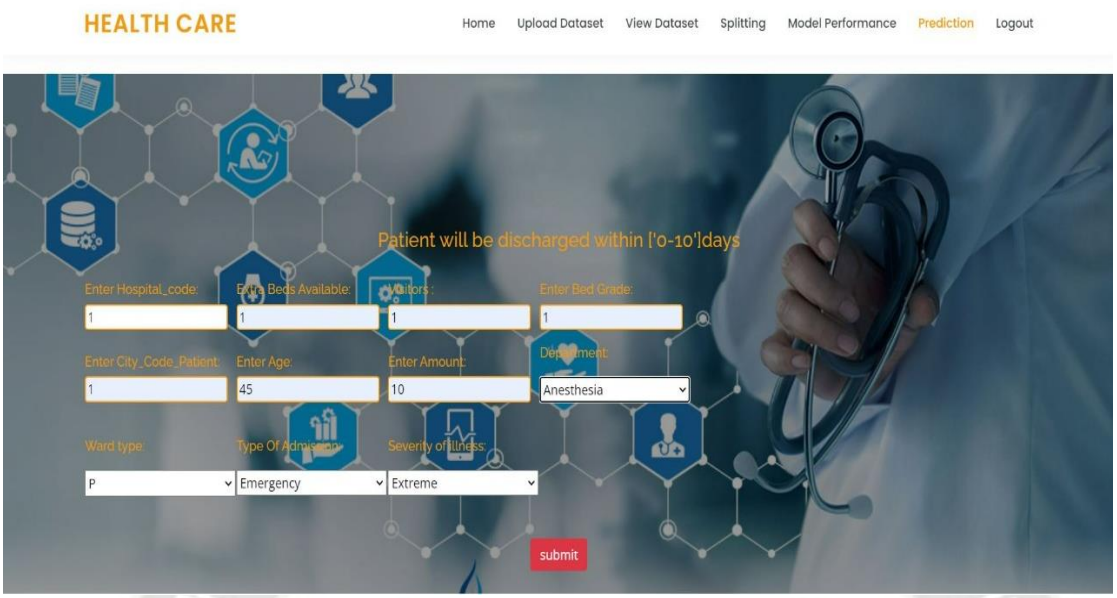


Fig. 12 Prediction of Extreme Severity of illness

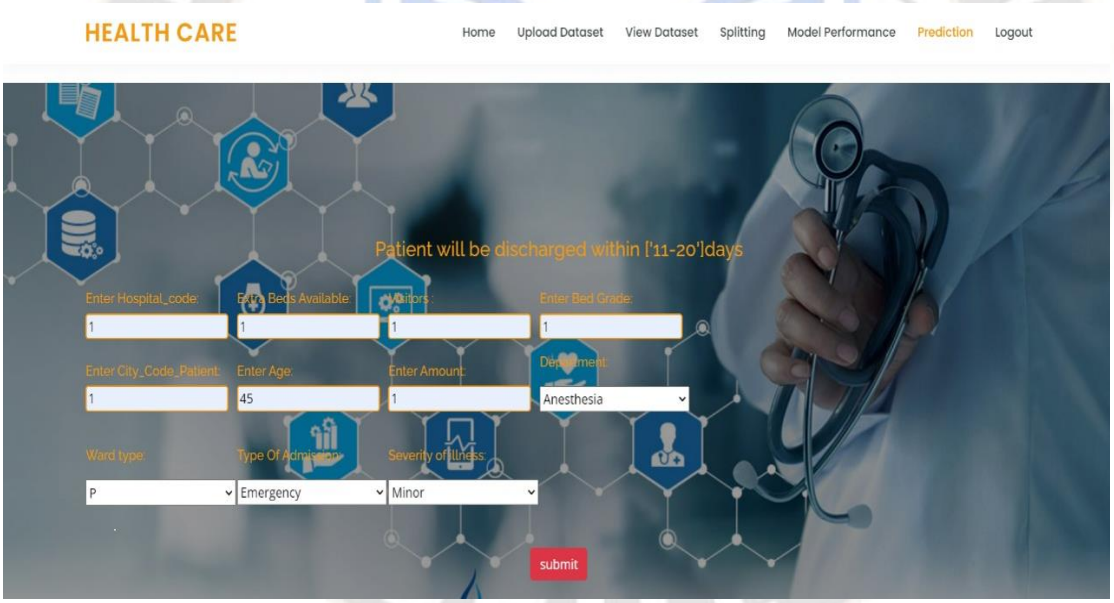


Fig. 13 Prediction of Minor Severity of illness

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5571
Article Received: 05 May 2022 Revised: 30 May 2022 Accepted: 20 June 2022 Publication: 31 July 2022

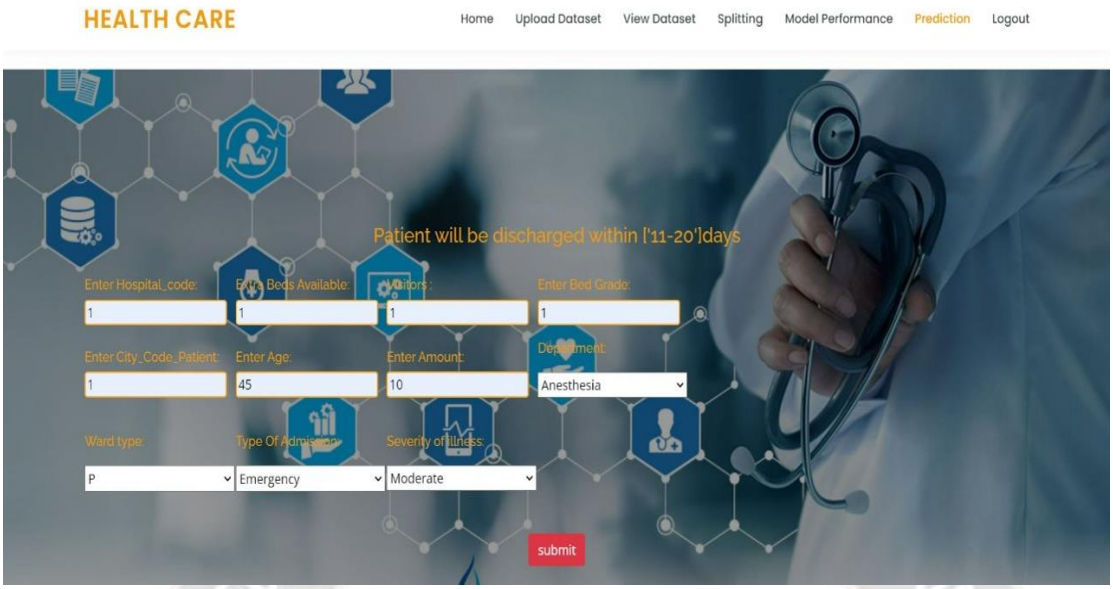


Fig. 14 Prediction of Moderate Severity of illness

6.3 Test Cases

The below table shows the test cases for the performance evaluation of our project.

Table 2. Initial Test Cases

Input	Output	Result
Input text	Tested for whether the input tells us in how many days a patient will be discharged	Success

6.4 Test cases Model building

The below table shows the test cases that is used for model building to evaluate the performance of our project.

Table 3. Test Cases Model building

S.NO	Test cases	I/O	Expected O/T	Actual O/T	P/F
1	Read the dataset.	Dataset.	Dataset need to read successfully.	Dataset fetched successfully.	P
2	Performing pre-processing on the dataset	Pre-processing part takes place	Pre-processing should be performed on dataset	Pre-processing successfully completed.	P
3	Model Building	Model Building for the clean data	Need to create model using required algorithms	Model Created Successfully.	P
4	Anaemia Estimation	Input fields provided.	Output should be number of days	Model has ability to predict in how many a patient will be discharged	P

When we investigate all the expected and actual outcomes tabulated in the above table, the outcomes were

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5571
***Article Received**: 05 May 2022 **Revised**: 30 May 2022 **Accepted**: 20 June 2022 **Publication**: 31 July 2022*

positive for all four cases to predict the rate of anaemia and thereby the discharge period when we evaluated our health sector serving application-based project.

7. Conclusion

We were able to devise a system for predicting the length of the days in which a diseased person would be get discharged after being hospitalized by investigating the rate of severity, admission type, and many more with our novel project that deployed three classifiers. We have considered the diseased persons who were hospitalized because of anaemia to predict the approximate discharge period. The consumer after filling the known data would get the approximate discharge period so that the diseased person need not stay in the hospital unnecessarily.

8. Future Scope of the Project

We aim to make use of various ailments in addition to anaemia in our application-based project. We also plan to examine prediction project with the more comprehensive public dataset to put forth a more reliable and accurate prediction model.

References

[1] Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., &Ghassemi, M. J. A. R. o. B. D. S. (2021). Ethical machine learning in healthcare. *4*, 123-144.

[2] Cuadrado, D., Riaño, D., Gómez, J., Bodí, M., Sirgo, G., Esteban, F., . . . Rodríguez, A. (2019). *Pursuing optimal prediction of discharge time in ICUS with machine learning methods.* Paper presented at the Conference on Artificial Intelligence in Medicine in Europe.

[3] Garg, D. K. . (2022). Understanding the Purpose of Object Detection, Models to Detect Objects, Application Use and Benefits. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(2), 01–04. https://doi.org/10.17762/ijfrcsce.v8i2.2066

[4] Ferdous, M., Debnath, J., & Chakraborty, N. R. (2020). *Machine learning algorithms in healthcare: A literature survey.* Paper presented at the 2020 11th International conference on computing, communication and networking technologies (ICCCNT).

[5] He, L., Madathil, S. C., Oberoi, A., Servis, G., Khasawneh, M. T. J. C., & Engineering, I. (2019). A systematic review of research design and modeling techniques in inpatient bed management. *127*, 451-466.

[6] Vanitha, D. D. . (2022). Comparative Analysis of Power switches MOFET and IGBT Used in Power Applications. International Journal on Recent Technologies in Mechanical and Electrical Engineering, 9(5), 01–09. https://doi.org/10.17762/ijrmee.v9i5.368

[7] Jabbar, M., Samreen, S., Aluvalu, R. J. I. J. o. E., & Technology. (2018). The future of health care: Machine learning. *7*(4.6), 23-25.

[8] Jain, V., & Chatterjee, J. M. J. C. S. (2020). Machine learning with health care perspective. 1-415.

[9] S. Mandal, S. Acharya, and R. Basak, "Solving Arithmetic Word Problems Using Natural Language Processing and Rule-Based Classification", Int J Intell Syst Appl Eng, vol. 10, no. 1, pp. 87–97, Mar. 2022.

[10] Kaur, P., Sharma, M., & Mittal, M. J. P. c. s. (2018). Big data and machine learning based secure healthcare framework. *132*, 1049-1059.

[11] Mustafa, A., & Rahimi Azghadi, M. J. C. (2021). Automated machine learning for healthcare and clinical notes analysis. *10*(2), 24.

[12] Pallathadka, H., Mustafa, M., Sanchez, D. T., Sajja, G. S., Gour, S., &Naved, M. J. M. T. P. (2021). Impact of machine learning on management, healthcare and agriculture.

[13] Siddique, S., & Chow, J. C. J. E. (2021). Machine learning in healthcare communication. *1*(1), 220-239.

[14] Vellido, A. J. N. c., & applications. (2020). The importance of interpretability and visualization in machine learning for applications in medicine and health care. *32*(24), 18069-18083.