JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Clinic Crowding Prediction Using SimPy Simulation and Logistic Regression

[1]Layla Hamad, [2] Layan Mohammed , [3]Rinad Misfer, [4] Ahmad M. Alkheder Almasabi

[123]Undergraduate Students, Department of Computer Science
[4] Faculty Member, Department of Computer Science
College of Computer Science, Najran University, Saudi Arabia

Abstract : This paper presents a combined simulation and machine learning approach to predicting clinic crowding. A discrete-event simulation was developed using SimPy to model patient arrivals, waiting times, and doctor availability. The simulation-generated dataset was used to train a Logistic Regression classifier, which achieved perfect accuracy. The results demonstrate the effectiveness of integrating simulation with machine learning for analyzing clinic performance.

IndexTerms **– Clinic Crowding, Simulation, SimPy, Machine Learning, Logistic Regression, Healthcare Modeling .**

I. Introduction

Managing clinic crowding is important for improving patient waiting times and resource utilization. This paper presents a simulation-based approach to modeling clinic behavior. The data generated from the simulation is used to train a machine learning model to classify clinic states as normal or crowded.

II. Simulation Model

A discrete-event simulation was created using SimPy. The model includes:

- 2 doctors
- Patient arrival every 5 minutes
- Service time between 5 and 15 minutes
- Total simulation time: 200 minutes

Each patient waits if doctors are busy, then receives service. The simulation records waiting time and service time for every patient.

III. Dataset generation

The simulation generates a dataset containing waiting_time, service_time, number of doctors, and arrival rate. A binary label is assigned:

- 0 = normal
- 1 = crowded

A case is labeled crowded when its waiting time exceeds the overall average waiting time.

IV. Machine learning model

A Logistic Regression classifier was trained using four features: waiting time, service time, number of doctors, and arrival rate. The dataset was split into 70% training and 30% testing.

Results:
Accuracy: 1.0
Precision: 1.0
Recall: 1.0
F1 Score: 1.0
The confusion matrix showed perfect classification performance.

V. Discussion

The model performed perfectly due to the structured nature of simulation data. While real clinic data may include noise, this experiment shows that combining simulation and machine learning is effective for analyzing crowding patterns.

VI. Conclusion

This study demonstrates a successful integration of simulation and machine learning to predict clinic crowding. The simulation produced a clear dataset, and Logistic Regression achieved perfect accuracy. This approach can be expanded to more complex models in future work.

References

[1] SimPy Documentation: https://simpy.readthedocs.io.

[2] Scikit-learn Documentation: https://scikit-learn.org.

[3] Pedregosa et al., “Scikit-learn: Machine Learning in Python,” JMLR, 2011.