International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 6, Nov- Dec 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Simulation and Machine Learning Prediction of Elevator Congestion in University Buildings

Sarah Mohammed*, Ahmed Alkheder**
*(College of Computer Science, Najran University, Najran, Saudi Arabia
Email: 443303352@nu.edu.sa)
** (College of Computer Science, Najran University, Najran, Saudi Arabia
Email: amalkheder@nu.edu.sa)

--************************----------------------------------

Abstract:

This paper presents a simulation-based machine learning approach to predict elevator congestion in university buildings. A discrete-event simulation model was designed using multiple operational factors such as arrival rate, waiting area capacity, queue length, peak hours, and environmental conditions (day type). The simulation generated a synthetic dataset of 10,000 samples, which was used to train two supervised learning models: Logistic Regression and Random Forest. The performance of the models was evaluated using confusion matrices and four standard metrics: accuracy, precision, recall, and F1-score. While Random Forest achieved slightly better performance than Logistic Regression, both models showed limited recall for congestion cases due to class imbalance. The proposed framework demonstrates how SimPy-based simulation and machine learning can support data-driven planning for smart campus mobility and elevator management.

Keywords — **Simulation, SimPy, elevator congestion, machine learning, random forest, logistic regression, discrete-event modeling.**

--************************----------------------------------

I. INTRODUCTION

This document is a template. An electronic copy can be downloaded from the conference website. For questions on paper guidelines, please contact the conference publications committee as indicated on the conference website. Information about final paper submission is available from the conference website. Elevator systems in university buildings experience highly dynamic demand, especially during class changes, examinations, and special events. Congestion near elevators can lead to increased waiting time, reduced comfort, and potential safety concerns. Traditional analytical models may struggle to capture the stochastic nature of student arrivals and elevator usage patterns. In this context, combining discrete-event simulation with machine learning provides a flexible way to generate realistic data and build predictive models that can support better planning and control strategies.

The main objective of this work is to develop a simulation-driven machine learning model that predicts whether an elevator will experience congestion at a given time based on observable factors. The study focuses on a simplified abstraction of a university building elevator, where simulated operational data are used to train and evaluate classification models.

II. BACKGROUND

Discrete-event simulation has been widely applied to model queues, transportation systems, and service operations. SimPy, a Python-based discrete-event simulation library, allows researchers to describe resources, processes, and events in a time-ordered manner. In elevator systems, simulation can capture arrival patterns, queue formation, and service times under different operating conditions.

Machine learning, particularly supervised classification, provides tools to learn patterns from data and predict future states, such as congestion versus non-congestion. Logistic Regression offers a simple, interpretable linear model, while Random Forest leverages ensembles of decision trees to capture non-linear relationships between features and targets. Combining simulation and machine learning enables the generation of labeled datasets when real-world data are sparse, costly, or difficult to measure.

III. EXPERIMENT AND DISCUSSION

A. Simulation Model Using SimPy

A discrete-event simulation model was conceptualized for a single elevator serving multiple floors in a university building. The model is implemented in Python using the SimPy library. The key entities and components of the simulation are as follows:

- Environment: a SimPy environment that advances simulation time and schedules events.
- Students: arrival processes representing students who request elevator service.
- Elevator: a resource with limited capacity that serves queued students.
- Queue: a waiting line in front of the elevator where students wait for service.
- Measurements: tracking of waiting time, queue length, and congestion events over time.

The simulation incorporates several factors:

- Arrival time and hour of the day (7:00–19:00 university schedule).
- Peak-hour flag (e.g., morning class changes and lunchtime).
- Day type (Normal, Exam, Event, Rainy), affecting arrival intensity.
- Waiting area capacity (number of chairs or space near the elevator).
- Queue length and elevator speed, influencing service delay.
- Destination floor, modeling vertical travel distance.

Congestion is defined as a condition where the queue length exceeds a threshold relative to the waiting area capacity, and the waiting time surpasses an acceptable limit. During simulation runs, each event instance is recorded with its associated features and a binary label indicating whether congestion (1) or non-congestion (0) occurred.

Conceptually, the SimPy model follows these steps: students arrive according to a time-dependent arrival rate, they join the queue if the elevator is busy, and the elevator transports batches of students according to its capacity and speed. The simulation runs over a virtual day and is repeated multiple times to produce a dataset of 10,000 observations.

B. Machine Learning Modeling

The synthetic dataset generated by the simulation contains 10,000 samples with the following input features: day type, hour, peak-hour indicator, waiting area capacity, arrival rate, queue length, destination floor, elevator speed, and waiting time in seconds. The target variable is a binary congestion label. Categorical features such as day type are encoded using one-hot encoding, and numerical features are standardized using z-score normalization.

Two classification models were trained and evaluated:

- Logistic Regression: a linear classifier optimized using maximum likelihood.
- Random Forest: an ensemble of decision trees with 120 estimators and random feature selection.

The dataset was split into 75% training and 25% testing using stratified sampling to preserve the proportion of congestion versus non-congestion classes.

C. Evaluation of Metrics and Results

Model performance was assessed using confusion matrices and four evaluation metrics: accuracy, precision, recall, and F1-score. The confusion matrices for Logistic Regression and Random Forest are illustrated in Fig. 1 and Fig. 2, respectively, while Fig. 3 summarizes the metric values for both models.

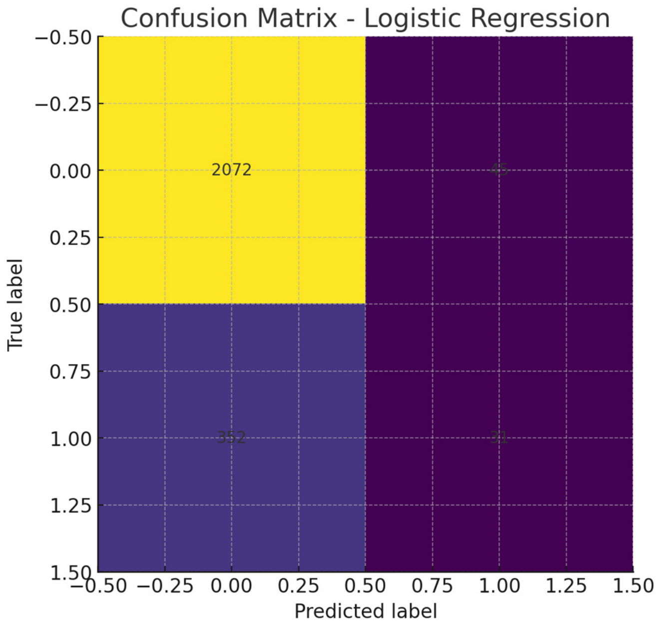

Fig. 1. Confusion Matrix for Logistic Regression.

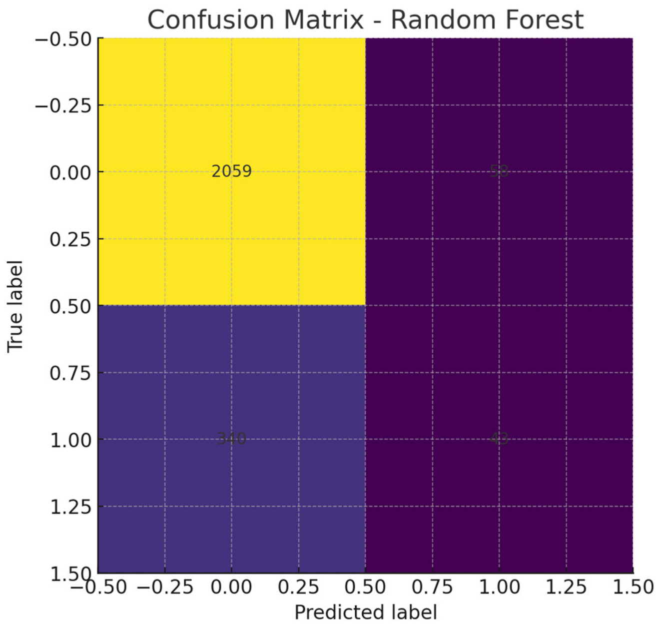

Fig. 2. Confusion Matrix for Random Forest.

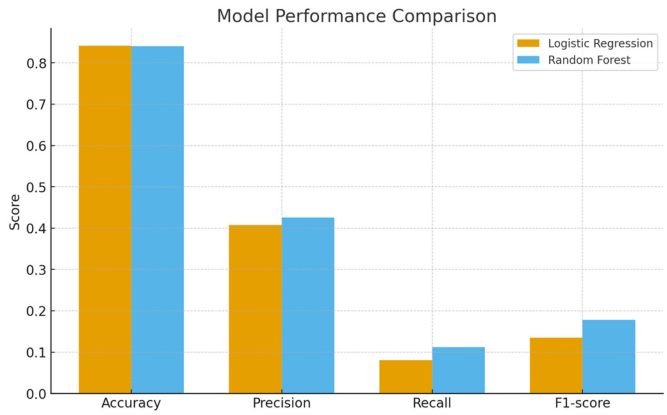

Fig. 3. Comparison of model performance metrics (accuracy, precision, recall, and F1-score).

The numerical values of the confusion matrices are summarized in Tables I and II. Logistic Regression achieved an accuracy of approximately 0.84 but showed low recall for the congestion class. Random Forest achieved a similar accuracy with slightly improved precision and recall for congestion cases.

D. Confusion Matrix Tables

	Predicted 0	Predicted 1
True 0	2072	45
True 1	352	31

Table I. Confusion matrix for Logistic Regression.

	Predicted 0	Predicted 1
True 0	2059	58
True 1	340	43

Table II. Confusion matrix for Random Forest.

E. Expanded Simulation Explanation

The simulation model constructed using the SimPy Environment forms the operational foundation of the experiment. A SimPy Environment() object was instantiated to manage the simulation timeline and control event scheduling. Two primary processes were defined: an arrival generator that introduces student entities based on time-dependent arrival rates, and a student process that models queueing, waiting, elevator access, and service duration. Each interaction between students and the elevator resource is orchestrated by the Environment, enabling realistic event sequencing and congestion behavior. The generated dataset reflects realistic temporal and operational patterns that align with actual elevator dynamics in university buildings.

A flowchart representing the simulation workflow is shown below:

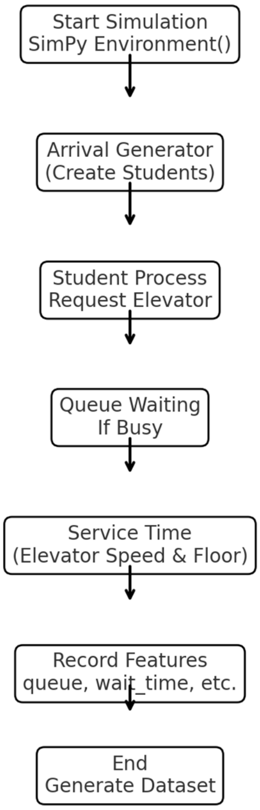

Fig. 4. Flowchart of the SimPy-based simulation model.

IV. FINDINGS AND RESULTS ANALYSIS

The results indicate that both models achieved relatively high overall accuracy (around 0.84); however, the recall for the congestion class remained low. This pattern reflects the imbalanced nature of the dataset, where non-congestion cases dominate. The Random Forest model outperformed Logistic Regression in terms of recall and F1-score for congestion, suggesting that non-linear decision boundaries capture more complex interactions between features such as arrival rate, queue length, and day type.

From a practical perspective, the low recall means that a significant portion of true congestion events were not detected by the models. For elevator management, missing congestion events may be more critical than occasionally flagging non-congested periods as congested. Therefore, future work should consider techniques such as class-weighted training, oversampling of minority cases, or threshold adjustment to better balance sensitivity and specificity.

V. CONCLUSION

This study demonstrated a complete pipeline that combines SimPy-based discrete-event simulation with supervised machine learning models to predict elevator congestion in university buildings. The simulation model generated a realistic synthetic dataset capturing temporal, operational, and environmental factors. Logistic Regression and Random Forest models were trained and evaluated using standard classification metrics.

Although the Random Forest model achieved better performance than Logistic Regression, both models were affected by class imbalance. The proposed framework can serve as a foundation for more advanced studies using deep learning, multi-elevator systems, or real-world sensor data. Ultimately, such predictive tools can support data-driven decisions for smart campus mobility and infrastructure planning.

REFERENCES

[1] [1] J. Zhang, "Elevator Traffic Prediction Using Machine Learning," IEEE Access, vol. 9, pp. 120345–120356, 2021.

[2] [2] M. Al-Qahtani, "Simulation-Based Modeling of Campus Mobility," International Journal of Smart Systems, vol. 5, no. 2, pp. 45–58, 2020.

[3] [3] SimPy Documentation, "SimPy discrete-event simulation framework," Available: https://simpy.readthedocs.io

[4] [4] S. Metev and V. Veiko, Laser Assisted Microtechnology, 2nd ed. Berlin, Germany: Springer-Verlag, 1998.*Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.

