Graduate Students, Information and Communications Technologies (ICT)

A STATION-LEVEL METRO RIDERSHIP PREDICTION MODEL CONSIDERING BUILT ENVIRONMENT; MACHINE LEARNING APPROACH

Aya Hasan AlKhereibi, Tadesse G. Wakjira, Murat Kucukvar, Uvais Qidwai, Deepti Muley, Abdulla Siam

INTRODUCTION

1. INTRODUCTION

RESEARCH PROBLEM

- Predicting metro ridership is an essential requirement for efficient metro operation and management.
- The dependence of metro ridership on the land use densities entails a need for an accurate predictive model.
- The state of Qatar recently introduced the metro and a new mode of transport, which required a sufficient planning tool for its optimal use.

AIM & OBJECTIVES

- The main aim of this research is to develop a novel machine learning (ML) based model to predict the metro station ridership utilizing the land use densities in the vicinity of metro stations.

RESEARCH SIGNIFICANCE

- Propose a novel ensemble ML model that predicts short-term metro ridership based on land use densities.
- Investigate the importance of built environment different key parameters in determining the metro ridership.
- Apply the ensemble ML models to a unique case of the State of Qatar.
- Propose applicable model for urban planners and transportation operators for optimal land use allocation to maximize metro ridership

METHODOLOGY

2. METHODS & MATERIALS

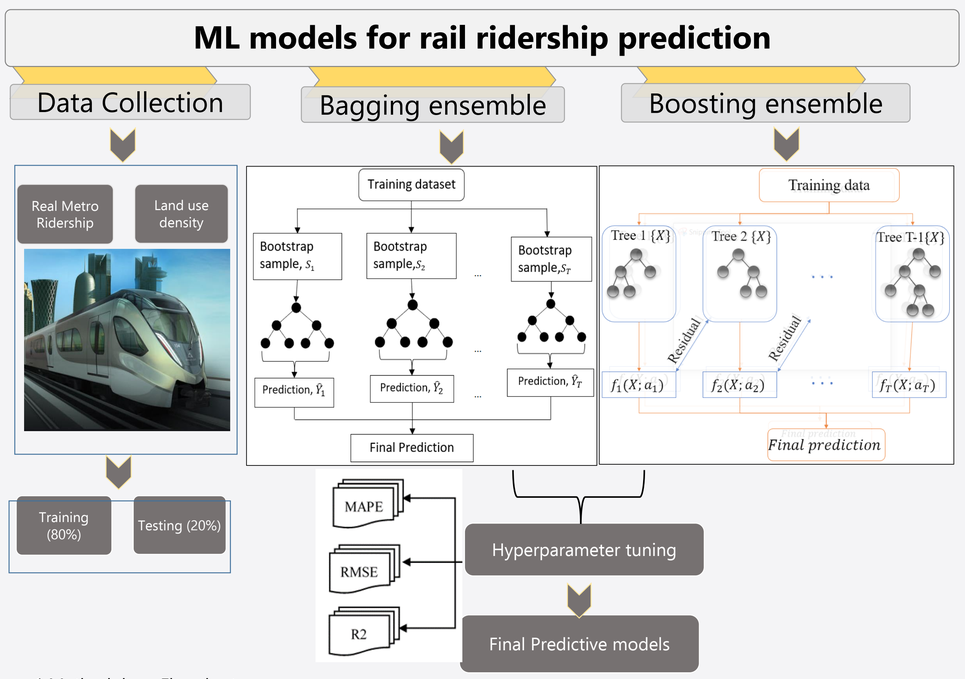


Figure 1 Methodology Flowchart

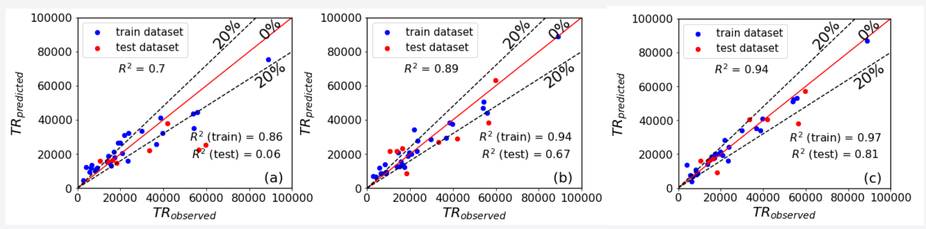


Figure 2 Comparison of predicted and observed total ridership (TR) using (a) random forest, (b) extremely randomized trees, and (c) gradient tree boosting models

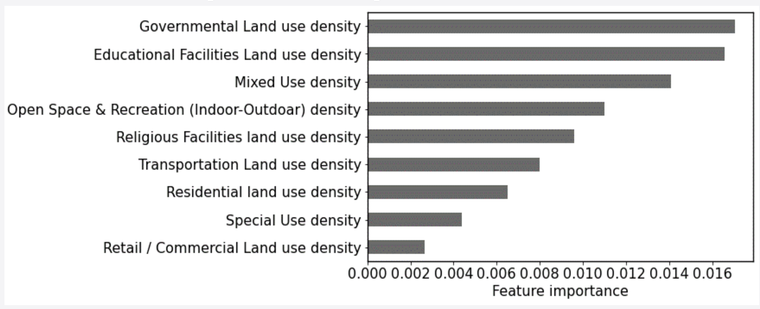


Figure 3 Feature importance analysis based on gradient tree boosting

RESULTS

3. KEY FINDINGS

Given the investigated features, the predictive capabilities of the models in terms of the MAPE, RMSE, and R^2 demonstrate that:

- The **Gradient Tree Boosting** model has the highest prediction capability,
- Followed by **Extremely Randomized Trees**. Among all models,
- **Random Forest** showed the least predictive performance.

Table 1 Models Predictive Performance

Models	Training dataset			Test dataset			Complete dataset		
	MAPE	RMSE	R^2	MAPE	RMSE	R^2	MAPE	RMSE	R^2
RF	0.39	7433.23	0.86	0.31	17747.60	0.06	0.37	10628.71	0.70
ERT	0.27	4731.97	0.94	0.43	10535.27	0.67	0.31	6486.95	0.89
GTB	0.29	3524.44	0.97	0.23	8067.18	0.81	0.28	4911.56	0.94

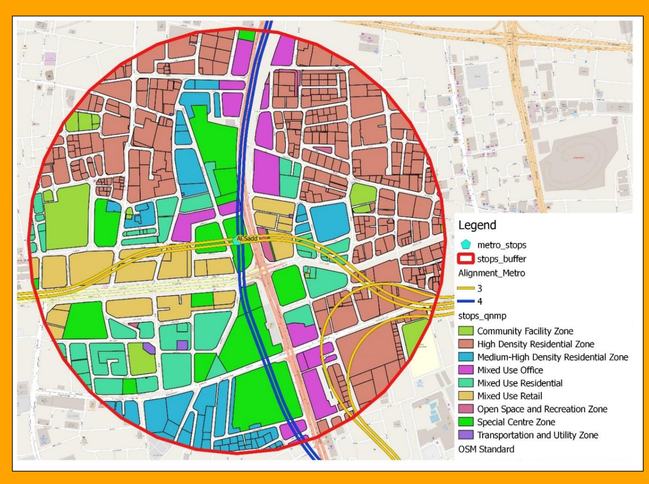
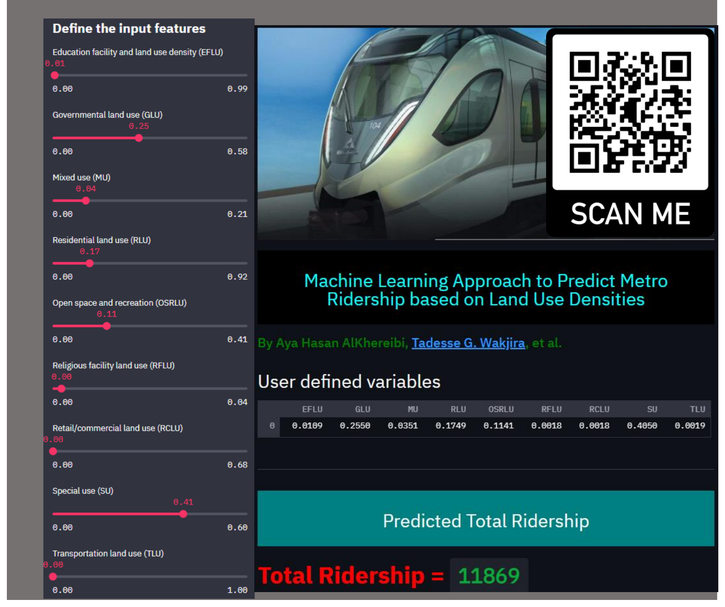


Figure 4 Al Sadd Metro Station with Land Use Densities in 800m catchment area

WEB-BASED APPLICATION



conclusion

4. CONCLUSION & RECOMMENDATIONS

The proposed model can be used by planning and operation authorities in their processes:

- to plan the land use around metro stations,
- predict the transit demand from those plans in order to achieve the optimal use of the transit system.
- predict ridership for newly planned developments based on land use densities.

Moreover, the proposed web-based application can easily be used by planners and researchers to investigate the key factors in determining metro ridership.

Recommendations for the future work

future study is recommended to focus on special circumstances such as the weekends, hourly variation, and special events that affect metro ridership.