



# Exploring the Potentials of Open-Source Big Data and Machine Learning in Shared Mobility Fleet Utilization Prediction

Mohamed Abouelela<sup>1</sup> · Cheng Lyu<sup>1</sup> · Constantinos Antoniou<sup>1</sup>

Received: 11 May 2022 / Revised: 18 March 2023 / Accepted: 21 March 2023 / Published online: 6 April 2023  
© The Author(s) 2023

## Abstract

The urban transportation landscape has been rapidly growing and dynamically changing in recent years, supported by the advancement of information and communication technologies (ICT). One of the new mobility trends supported by ICT is shared mobility, which has a positive potential to reduce car use externalities. These systems' recent and sudden introduction was not adequately planned for, and their rapidly growing popularity was not expected, which resulted in the urgent need for different stakeholders' intervention to ensure efficient services' integration within the urban transportation networks and to grant an effective system operation. Several challenges face shared mobility, including fleet size management, vehicle distribution, demand balancing, and the definition of equitable prices. In this research, we developed a practical, straightforward methodology that utilizes big open-source data and different machine learning (ML) algorithms to predict the daily shared-e-scooter fleet utilization (the daily number of trips per vehicle) that could be used to drive the system's operation policies. We used four ML algorithms with different levels of complexity, namely; Linear Regression, Support Vector Regression, Gradient Boosting Machine, and Long Short-Term Memory Neural Network, to predict the fleet utilization in Louisville, Kentucky, using the knowledge the models get from the training data in Austin, Texas. The Gradient Boosting Machine (LightGBM) was the model with the best performance prediction based on the different evaluation measures. The most critical factors impacting daily fleet utilization prediction were temporal time series features, sociodemographics, meteorological data, and the built environment.

**Keywords** Shared mobility · Micromobility · Shared-E-scooter · Machine learning · Demand prediction · Open source data

## Introduction

The urban population is rapidly growing at an unexpected rate led by the increasing urbanization movement; the UN expects that by 2050, 80% of the world population will live in urban areas, compared to 49% in 2010 (United Nations Department of Economic and Social Affairs 2018). The urban population growth is coupled with a substantial increase in travel demand, air pollution, accidents, and

more congested urban networks (Zannat and Choudhury 2019). Investment in infrastructure and significant land-use changes might be required to meet the expected growth of travel demand; however, such solutions need significant investments and long time processes to materialize, which are not always viable solutions (Bhattacharya et al. 2012; Estache 2010). Innovative solutions supported by the latest advancement in information and communication technologies (ICT) could represent a smart, efficient way to cater to the increased demand. The advancement of ICT already supports (fully or partially) the revolutionizing of the urban transportation systems; in the main developing areas of transportation; electrification, sharing, and automation (Sperling 2018).

Shared mobility services are one example of the recent innovative solutions that could cater to the expected increase in travel demand. These services provide commuters with access to different vehicle types or the ability to share rides based on the users' needs (Shaheen et al. 2016; Shared and

---

✉ Mohamed Abouelela  
Mohamed.abouelela@tum.de

Cheng Lyu  
cheng.lyu@tum.de

Constantinos Antoniou  
c.antoniou@tum.de

<sup>1</sup> Chair of Transportation System Engineering, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany

Digital Mobility Committee 2018). The umbrella of shared mobility services covers different service types that can be split into two main categories; (i) sharing of rides based on different operational schemes such as the case of e-hailing, ridesharing, ride pooling, and alternative transit systems. (ii) The direct use of the vehicle based on need, such as the case of carsharing and micromobility services, e.g., bikesharing and shared-e-scooter, which are the focus of this research. Shared mobility services have many positive potential impacts on the urban environment, including reducing vehicular traffic (Aboulela et al. 2021b, 2023), reducing energy consumption, and increasing transport system efficiency by achieving saving in travel time and travel costs (Becker et al. 2020). Notwithstanding the possible positive effects of shared mobility services, some of them have integration, planning, and policies challenges following their sudden and novel introduction to the urban environment, such as the case of shared-e-scooters, we will refer it as *scooters* in the rest of the article. Scooters are one of the youngest members of the shared mobility services family, launched in July 2017 by Lime ([www.li.me](http://www.li.me)) in Santa Monica, California. 38.5 million trips were completed by scooters in the USA by the end of 2018, representing 45.8% of the total micromobility trips for the same year. In the following year, 2019, the number of scooter trips raised to 88.5 million in 109 cities in the USA, representing an exponential increase of 130% of the previous year's trips (NACTO 2019). Scooter use and growth were not limited to the US, but it was a global phenomenon observed in Asia, Europe, and Australia (Santacreu et al. 2020; Heineke et al. 2019; Møller and Simlett 2020). The micromobility market is expected to grow to between \$330 and \$500 billion by 2030 (Heineke et al. 2019). However, the growth of scooters faces several challenges, such as the increase in related injuries (Yang et al. 2020a; Namiri et al. 2020), defining the optimal fleet size, vehicles optimal redistribution strategies, speed limits enforcement, and equity regulations (Janssen et al. 2020). In order to further study these problems and define their causes and factors leading to them, more data is required.

Recently, the term big data has gained popularity and attracted more effort from the industry and the research sides to explore the opportunities it can create. The advancement of ICT has also opened the horizon for collecting and analyzing new types of data in large quantities, or so-called big data (Stojanović and Stojanović 2020; Chaniotakis et al. 2020). Other factors have helped in the collection of large amounts of data, such as but not limited to; the increase in the computational power of computers, the decreasing cost of data storage, and the exciting direction towards smart cities platforms, have also enriched the interests in big data (Iliashenko et al. 2021; Xin et al. 2020; Torre-Bastida et al. 2018; Zhu et al. 2018). Big data have been examined in many applications

related to transportation research, e.g., estimating transit network origin–destination flows (Liu et al. 2021a), availability of parking supply using sentiment analysis of the location-based social network data (LBSN) (Chaniotakis et al. 2022; Jiang and Mondschein 2021), improve traffic management and traffic planning (Haghighat et al. 2020), and the impact of pricing schemes changes on bikesharing use (Venigalla et al. 2020). Different entities, primarily operators and cities' authorities, are currently sharing their data (big based on volume, velocity, or variety) openly to encourage the innovation of new methods and ideas to improve the urban environment, to increase integration between the different transportation services, and to help in regulating and dynamically adjusting the operation of various shared mobility services within the urban environment (Durán-Rodas et al. 2020a; Iliashenko et al. 2021).

In this research, we use the publicly available scooter trips data from two American cities, Louisville, Kentucky, and Austin, Texas, in combination with other open data sources, discussed in detail in the methodology sections to explore the potential and accuracy of using open-source data and machine learning (ML) techniques to predict the scooter daily fleet utilization (number of trips per vehicle). The main goal of this research is to create and develop a framework that could help the different stakeholders involved in the operation, organization, and governance of the micromobility services to integrate the service in the urban environment efficiently and to facilitate the policy-making process.

The main contribution of this research comes from developing a framework for scooter fleet utilization prediction (daily number of trips per vehicle) using different sources of publicly available data and how this information is processed (including feature engineering) to obtain the optimum prediction results in terms of prediction error minimization.

The contribution of this work can be summarized as follows:

- Using open-source big data and ML techniques to predict the daily scooter fleet utilization (daily number of trips per vehicle).
- Compare the prediction results using different ML techniques; Gradient boosting decision tree (GBDT), Linear regression (LR), Support Vector Regression (SVR), and more complex deep learning techniques such as Long Short-Term Memory Neural Network (LSTM-NN).
- The prediction period was more extended than 1 year, which is longer than the periods used in most previous research, representing a long-term forecast horizon of operation.
- The prediction model showed accurate results when used to predict the test dataset, discussed in more detail in the following sections; therefore, it could be implemented in

real-life and the scooter deployment, organization, and governance processes.

- The proposed methodology is a practical yet simple<sup>1</sup> method to transfer the scooter fleet size utilization prediction learned from one city to another; which implies that same frame work could be used for other cities.
- The developed methodology, with its capabilities to be transferred to other cities, is to be used to define the daily,<sup>2</sup> weekly, or seasonal fleet size based on the predicted utilization rate for the existing services and the planned one, which could help in more efficient system operation by dynamically predicting and then deploying the number of vehicles needed to cater for the demand and not to deploy a fixed number of vehicles that do not account for the demand seasonality and fluctuation. In addition, deploying the vehicle based on the actual demand could reduce the number of ideal vehicles and facilitate vehicle re-balancing, distribution processes, and the subsequent vehicle kilometer traveled resulting from the distribution process.

The rest of this paper is organized as follows; “[Literature Review](#)” section discusses the current literature related to the discussed topic, “[Methods, Data, and Case Study](#)” section shows the data used and methodology, “[Analysis Results](#)” section shows the analysis for the collected data and estimated models. Finally, “[Discussion, and Conclusion](#)” section discusses the results and the conclusion of the research.

## Literature Review

The literature review is organized as follows; in the first part, we define the reasons to use shared mobility services, their potential positive impacts, factors impacting their demand, and challenges shared mobility faces. The second part summarizes the potential of new sources of data and different ML techniques used in shared mobility studies, focusing on shared vehicle systems (carsharing, bike sharing, and shared-E-scooters) and their potential for demand prediction compared to traditional regression models.

Shared mobility is a rapidly growing trend in recent years encouraged by many factors such as but not limited to, travel time savings, ease of payment, fare transparency, trip cost, comfort and security, or even health benefit such as in the case of bikesharing (Abouelela et al. 2022; Tirachini and del

Río 2019; Tirachini 2020; Tirachini and Gomez-Lobo 2020; Cerutti et al. 2019; Circella et al. 2018; Nikitas et al. 2015; Schaefers 2013), the popularity of smartphones and the development of mobile applications, or the general advancement of ICT (Spinney and Lin 2018; Schmöller et al. 2015). Moreover, shared mobility supports sustainability goals, or at least they can be described as more sustainable transport systems than private vehicle use; for example, the use of car-sharing systems could have the positive potential for reducing negative traffic externalities (Kostic et al. 2021). Also, shared mobility could reduce the vehicular kilometer traveled (VKT) as in the case of bikesharing, shared-e-scooter, and pooled rides (Ting et al. 2021; Abouelela et al. 2021b; Tirachini and Gomez-Lobo 2020; Ricci 2015). Integrating shared mobility services in the urban environment faces several challenges, mainly tied to the systems’ governance and management. These operational problems are more avid and critical for vehicle-sharing systems (scooter sharing, bike-sharing, and carsharing), especially for free-floating systems, compared to other forms of shared mobility. The main problems are; fleet size management, spatial and temporal demand prediction and estimation, fleet geographical distribution and re-distribution, deciding on the optimal pricing schemes, use equity, accessibility of the service, operational hours, and geographical limits (zonal fencing) (Duran-Rodas et al. 2020b; Turoñ et al. 2019; Liu et al. 2018; Ko et al. 2019; Shaheen and Cohen 2013; Weikl and Bogenberger 2013). It is to be noted from the previously mentioned challenges faced by the shared mobility services that most of the challenges are directly linked to the travel demand; therefore, understanding factors impacting the demand and demand prediction is a must to improve shared mobility operations and integration.

Several exogenous factors impact the use of shared mobility, and they can be categorized into four main groups. The first group is related to the shared vehicle’s systems (bike sharing, scooter, and carsharing), such as the presence of docking stations and vehicles availability (Reck et al. 2021; Raux et al. 2017; De Lorimier and El-Geneidy 2013). The second group of factors is the infrastructure-related factors, including the availability of bike lanes, the density of road intersections, and the availability of parking lots (Müller et al. 2017; Chen et al. 2018; Hu et al. 2018). Meteorological conditions also play a significant role in shared mobility use, which is evident in the case of bikesharing, carsharing, and scooters. In contrast, adverse weather conditions significantly reduce bike sharing and scooter sharing use; it increases the use of carsharing (Yoon et al. 2017; Lin et al. 2018; Shen et al. 2018; Abouelela et al. 2021b). The last group is the land use and built environment and points of interest (POI), where different land uses impact the various

<sup>1</sup> Refer to “[Methods, Data, and Case Study](#)” section, and “[Model Transfer](#)” section, where we explain the used transfer methodology namely label differencing and sample normalization.

<sup>2</sup> We predicted the fleet utilization rate daily so that it can be aggregated to courser time units, e.g., weekly, monthly, or season based.

services. For example, mixed land uses are associated with carsharing use, commercial land use<sup>3</sup> is linked to bikesharing, scooter sharing and carsharing use (Kim et al. 2015; Hu et al. 2018; Abouelela et al. 2021a). Also, POI impact the use of shared mobility, where educational institutes, schools, and universities, were found to be associated with the increased use of carsharing and bikesharing (El-Assi et al. 2017; Mattson and Godavarthy 2017; Sun et al. 2017; Kim et al. 2012; Sun et al. 2017).

The generation and availability of big data, sometimes publicly available, from new sources supported by the advancement of ICT, in addition to the advancement in processing and data-storing methods, has facilitated data use and the further development of applications. Several examples can be referred to, such as the use of big spatial data to evaluate the relation between housing rental and carsharing use in Korea (Choi and Yoon 2017), identifying scooter users' segments from trip data in Germany (Degele et al. 2018), predicting the demand for bikesharing systems using a combination of weather data and bike booking data from New York (Cantelmo et al. 2020), the use of smartphone applications to understand pedestrian route choice behavior (Sevtsuk et al. 2021), and using news report to investigate scooters' crashes (Yang et al. 2020a).

The capabilities of ML with the combination of big data use have already been explored in the research related to shared mobility use; for the different shared vehicle services. For example, shared micromobility, which is mainly praised for its potential to solve the last mile problem (Baek et al. 2021; Fearnley et al. 2020; Luo et al. 2021), where Yang et al. (2016) proposed a spatio-temporal mobility model of bike-sharing and present an OD demand (check-in and check-out demand) prediction mechanism based on historical bike-sharing and meteorological data. They used a probabilistic model for the check-in demand, while a random forest (RF) model was introduced for check-out demand. Factors such as time of the day, day of the week, holidays, and weather conditions were found to be significant in predicting demand. Gammelli et al. (2020) proposed a general method for censorship-aware demand modeling by devising a censored likelihood function; censorship-aware demand is used to simulate reality. Transport demand is highly dependent on supply for shared mobility services, where services are often limited. Predictive models would necessarily represent a biased version of the actual demand without explicitly accounting for the supply restriction. The censored likelihood within a Gaussian Process model was incorporated and validated the limiting effect of supply on

bike-sharing demand data to counter the previous problem. ML was also used in the case of scooter sharing, but not extensively as used for other modes; Saum et al. (2020) combined Box-Cox transformation, seasonal autoregressive moving average (SARIMA), and family of generalized autoregressive conditional heteroskedasticity (GARCH) models to predict hourly demand and volatility for scooter demand, for a limited period in; Thammasat University, Thailand. Deep learning models are also becoming more popular and widely used in transport. Gao and Lee (2019) propose a moment-based model with a new hybrid approach that combines a fuzzy C-means (FCM)-based genetic algorithm (GA) with a back propagation network (BPN) to predict bikesharing rentals. Xu et al. (2018) developed a long short-term memory (LSTM) model based on different data types (trip data, weather data, air quality data, and land use data) to predict the bikesharing trip generation and attraction for different time intervals (10, 15, 20, and 30 min). They also compared the model with other popular ML models, including one-step forecast, ARIMA, optimized gradient boosting algorithm (XGBoost), support vector machine (SVM), artificial neural network (ANN), and recurrent neural network (RNN).

Also using deep learning, Yang et al. (2020b) focused on graph features; they extracted time-lagged variables describing graph structures and flow interactions from bike usage data. These variables include graph node Out-strength, In-strength, Out-degree, In-degree, and PageRank. The results proved that different machine learning approaches (XGBoost, MLP, LSTM) improve the prediction accuracy when time-lagged graph information is included. Zhang et al. (2019) used a deep learning model to predict the hourly travel demand using an LSTM model and compared it with different ML algorithms such as support vector regression (SVR), autoregressive integrated moving average model (ARIMA) for carsharing systems. The results demonstrated that LSTM performs better in terms of performance and precision. Also, Luo et al. (2019) predicted dynamic demand based on graph features. The model was tested on real-world shared electric vehicle (EV) data, showing accurate prediction results. It is worth mentioning that, in comparison with traditional regression techniques, regression models generally show a poor prediction power when compared to ML algorithm; for example, in the case of carsharing, Müller et al. (2017) used a negative binomial statistical model to predict the vehicles demand, and the models' R-squared ( $\rho^2$ ) were around (0.07). Also, Younes et al. (2020) used negative binomial models to predict the average hourly trips for bikesharing and shared-e-scooter with ( $\rho^2$ ) ranging between (0.14–0.20). These examples show the poor prediction capabilities of regular regression models compared to ML, which supports the potential of using ML techniques for further research. Table 1 shows a summary of some selected studies and the used ML techniques, used performance evaluation matrices, and the recommended technique if applicable.

<sup>3</sup> According to the American Planning Association APA ([planning.org](http://planning.org)), commercial land use is the land use that contains commercial retail and wholesales, business offices; while mixed land use is the combination of more than one land use in the same area such as residential, public and semi-public, and parks and open spaces.

**Table 1** Selected examples of ML techniques used in selected studies for different shared vehicle systems from selected studies

References	ML technique	Performance evaluation metrics	Best performing ML*
<i>Micro-mobility</i>			
Yang et al. (2016)	Check-in: probability model for the undocked bikes; Check-out behaviour: RF	RMLSE	NA
Saum et al. (2020)	SARIMA + GARCH and their variation	MAE, MSE, MAPE	Modified SARIMA (BoxCox-SARIMA and SARIMA-PGARCH) outperformed other models
Yang et al. (2020b)	Feature extraction + XGBoost/MLP/ LSTM	MAPE, RMSE	LSTM
Gao and Lee (2019)	FCM-based GA with BPN	RMSE, MAE	NA
Xu et al. (2018)	LSTM; HA, ARIMA, XGBoost, SVM, ANN, RNN	RMSE	LSTM
<i>Car-sharing</i>			
Zhang et al. (2019)	LSTM; SVR, ARIMA, smoothing	MSE, $R^2$	LSTM
Luo et al. (2019)	Multi-graph Dynamic GCN	RMSE, Error Rate	NA

\* NA refers to no comparison between the different models was performed

## Methods, Data, and Case Study

### Methods

#### Problem Statement and Framework Overview

Figure 1 shows the overview of the proposed methodology framework. We employed in this research the model transfer problem for time series prediction to predict scooters' fleet utilization (Zhang et al. 2020). Given the historical demand data in the source city (Austin) alongside the pilot stage<sup>4</sup> demand data in the target city, Louisville, a time series model was trained and applied to predict the future fleet utilization in the target city. The source city is the city that provides us with the long-term patterns of historical demand and fleet utilization changes, whereas the target city only has information on demand changes over a short pilot stage.

The historical data of a city is denoted as  $\mathbf{D} = \{\mathbf{d}_i\}_{i=1}^Z$ , where  $Z$  is number of census tracts (demand aggregation zones).  $\mathbf{d}_c = \{\mathbf{t}_c, \mathbf{z}_c\}$ <sup>5</sup> is the data of census tract  $c$ , consisting of both historical time series  $\mathbf{t}_c \in \mathbb{R}^L$  and auxiliary census tract attributes  $\mathbf{z}_c \in \mathbb{R}^N$ , where  $L$  is the length of the time series and  $N$  is the length of auxiliary attributes. The data of the source city and the target city can be respectively

denoted by  $\mathbf{D}_S$  and  $\mathbf{D}_T$ . The two lengths  $L$  and  $N$  can be determined based on the richness of data rather than fixed. For example, longer pilot stage duration and more accessible land use attributes allow the choice of larger  $L$  and  $N$ .

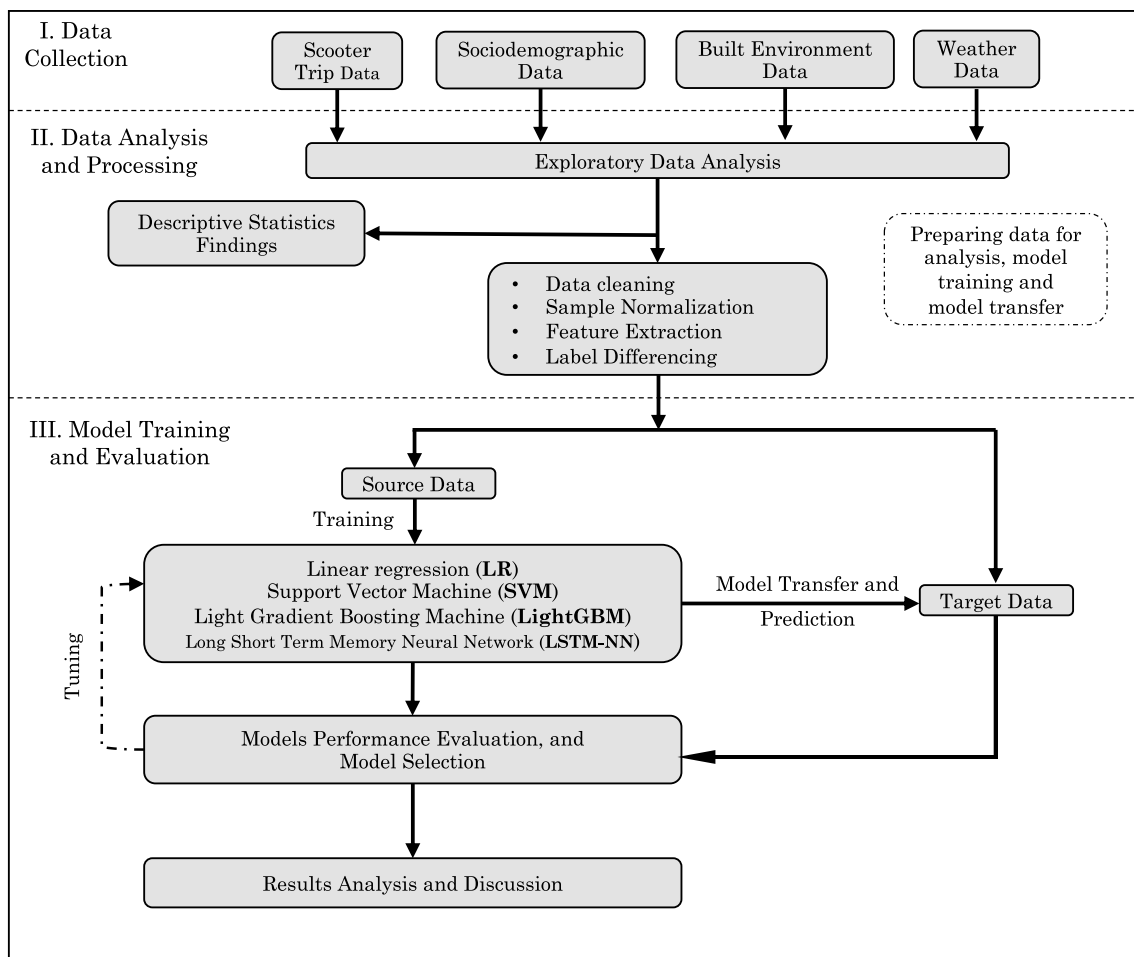
An autoregressive formulation was adopted for the time series prediction problem, such that it was transformed into a supervised ML problem. The raw data was split into two samples for model training and testing. A sample is described by a vector pair  $(\mathbf{x}_i, y_i)$ , where  $i$  is the index of the sample. The first element  $\mathbf{x}_i = \{\mathbf{x}'_i\}_{j=1}^m \in \mathcal{X}$  is an  $m$ -dimensional feature vector, which is comprised of  $m$  features extracted through feature engineering from the census tract attributes and the time series data of  $w$  consecutive days in a specific census tract  $c$ , i.e.,  $\mathbf{t}_c^{(i:i+w)}$ . The label  $y_i \in \mathcal{Y}$  is the succeeding time series value in census tract  $c$ , i.e.,  $t_c^{(i+w)}$ .

The ordinary time series prediction problem aims at learning an accurate mapping  $f : \mathcal{X} \rightarrow \mathcal{Y}$  on future time steps in the same time series as in  $\mathbf{D}_S$ . However, the model transfer of the time series prediction problem aims at learning another mapping  $f' : \mathcal{X} \rightarrow \mathcal{Y}$  from  $\mathbf{D}_S$ , but still performs well on the time series of  $\mathbf{D}_T$ . The foremost difficulty in model transfer lies in the inconsistency between the distributions of data in  $\mathbf{D}_S$  and  $\mathbf{D}_T$ , also known as the covariate shift. To address this problem, we proposed a simple yet effective approach to align the distributions of time series in two cities and minimize the generalization error of the time series prediction model. Following the common ML procedures, the four-step pipeline of (*sample construction—feature engineering—model training—inference*) was adopted. Two strategies were used to facilitate the transfer of the time series prediction model, namely the sample normalization and the label difference. Note that the proposed framework

<sup>4</sup> We considered the first 3 months the service deployed in the source city as a pilot stage, as commuters are generally trying to get familiar with the service, and it is the same period used by other cities to evaluate scooters' deployment such as Minneapolis, MN (Abouelela et al. 2023)

<sup>5</sup> The length of  $\mathbf{t}_c$  depend on the available amount of historical time series data, and  $\mathbf{z}_c$  depends on the other auxiliary variables length.





**Fig. 1** The used methodological framework

is compatible with various base ML models, which will be discussed in “[Base Models](#)” section.

### Feature Engineering

Feature engineering is an important step for ML model development. Raw data were examined and processed to predict important information before using it in the modeling process. Our model incorporated two categories of features, namely time-series features and auxiliary features. Historical time series characteristics were included in the feature set so that the model could learn the patterns of time series dynamics from them. Liu et al. (2020), Zhang et al. (2016) recommend considering; (i) neighboring information, (ii) periodical information, and (iii) trend information for accurate time series prediction. Neighboring information contains the demand values on the neighboring days of the target prediction day, informing the model of the recent level of demand. In this study, the number of neighbors is set as five. Periodical

information contains historical demand values on a weekly basis, as repetitive weekly peaks can be observed from Fig. 3. Trend and seasonality needed to be removed through differencing before applying classical time series prediction tools like ARIMA (Kwiatkowski et al. 1992). Although ML models do not explicitly assume stationarity for time series prediction, a nonstationary time series is not always suitable for prediction without preprocessing, especially for decision-tree-based models, which is explained in the following section. Therefore, a first-order differencing<sup>6</sup> was applied to the demand data as prediction labels. Table 2 shows the summary of the statistics of the time series features. Apart from time-series features, auxiliary information was proven to be of great help in prediction tasks (Liu et al. 2021b; Lyu et al. 2020; Wessel

<sup>6</sup> First-order differencing refers to computing the difference between two consecutive values in the time series. Denote the  $i$ -th observation of a time series as  $x_i$ , the transformed label  $y_i$  will be defined as follows;  $y_i = x_i - x_{i-1}$ .

**Table 2** Time series feature summary statistics

Feature	Type <sup>a</sup>	Austin		Louisville	
		Mean	StD	Mean	StD
Demand (previous day) ( $\times 10^{-3}$ )	N	6.36	33.91	9.67	53.24
Demand (2 days ago) ( $\times 10^{-3}$ )	N	6.37	33.94	9.69	53.30
Demand (3 days ago) ( $\times 10^{-3}$ )	N	6.37	33.96	9.71	53.36
Demand (4 days ago) ( $\times 10^{-3}$ )	N	6.38	33.99	9.73	53.41
Demand (5 days ago) ( $\times 10^{-3}$ )	N	6.38	34.00	9.74	53.47
Demand (1 week ago) ( $\times 10^{-3}$ )	P	6.39	34.03	9.78	53.59
Demand (2 weeks ago) ( $\times 10^{-3}$ )	P	6.43	34.18	9.92	54.00
Demand (3 weeks ago) ( $\times 10^{-3}$ )	P	6.48	34.35	10.06	54.42
Demand (4 weeks ago) ( $\times 10^{-3}$ )	P	6.52	34.51	10.21	54.85
Demand (difference between 1 and 2 days ago) ( $\times 10^{-3}$ )	T/N	0.00	12.18	-0.02	24.89
Demand (difference between 2 and 3 days ago) ( $\times 10^{-3}$ )	T/N	0.00	12.19	-0.02	24.91
Demand (difference between 3 and 4 days ago) ( $\times 10^{-3}$ )	T/N	0.01	12.20	-0.02	24.94
Demand (difference between 4 and 5 days ago) ( $\times 10^{-3}$ )	T/N	0.01	12.21	-0.02	24.97
Demand (difference between 5 and 6 days ago) ( $\times 10^{-3}$ )	T/N	0.01	12.22	-0.02	25.00
Demand (difference between 6 and 7 days ago) ( $\times 10^{-3}$ )	T/N	0.01	12.23	-0.02	25.02
Demand (difference between 8 and 9 days ago) ( $\times 10^{-3}$ )	T/P	0.00	12.24	-0.02	25.08
Demand (difference between 15 and 16 days ago) ( $\times 10^{-3}$ )	T/P	0.00	12.29	-0.02	25.28
Demand (difference between 22 and 23 days ago) ( $\times 10^{-3}$ )	T/P	0.00	12.35	-0.02	25.48
Demand (difference between 29 and 30 days ago) ( $\times 10^{-3}$ )	T/P	0.01	12.41	-0.02	25.69
Demand (average of the past week) ( $\times 10^{-3}$ )	T	6.40	31.99	9.72	49.04
Demand (variance of the past week) ( $\times 10^{-3}$ )	T	0.16	3.10	0.52	6.24
Demand (range of the past week) ( $\times 10^{-3}$ )	T	6.49	32.93	13.44	61.36
Citywide demand (previous day) ( $\times 10^{-2}$ )	N	76.95	36.96	80.28	54.56
Citywide demand (2 days ago) ( $\times 10^{-2}$ )	N	77.02	36.96	80.42	54.53
Citywide demand (3 days ago) ( $\times 10^{-2}$ )	N	77.10	36.95	80.58	54.49
Citywide demand (4 days ago) ( $\times 10^{-2}$ )	N	77.16	36.95	80.73	54.46
Citywide demand (5 days ago) ( $\times 10^{-2}$ )	N	77.21	36.96	80.88	54.42
Citywide demand (1 week ago) ( $\times 10^{-2}$ )	P	77.29	37.00	81.21	54.32
Citywide demand (2 weeks ago) ( $\times 10^{-2}$ )	P	77.78	36.93	82.37	53.98
Citywide demand (3 weeks ago) ( $\times 10^{-2}$ )	P	78.37	36.77	83.49	53.70
Citywide demand (4 weeks ago) ( $\times 10^{-2}$ )	P	78.92	36.67	84.72	53.27
Citywide demand (difference between 1 and 2 days ago) ( $\times 10^{-2}$ )	T/N	0.06	25.50	-0.16	33.54
Citywide demand (difference between 2 and 3 days ago) ( $\times 10^{-2}$ )	T/N	0.06	25.52	-0.16	33.57
Citywide demand (difference between 3 and 4 days ago) ( $\times 10^{-2}$ )	T/N	0.06	25.54	-0.16	33.61
Citywide demand (difference between 4 and 5 days ago) ( $\times 10^{-2}$ )	T/N	0.08	25.56	-0.16	33.65
Citywide demand (difference between 5 and 6 days ago) ( $\times 10^{-2}$ )	T/N	0.09	25.58	-0.17	33.68
Citywide demand (difference between 6 and 7 days ago) ( $\times 10^{-2}$ )	T/N	0.09	25.60	-0.17	33.72
Citywide demand (difference between 8 and 9 days ago) ( $\times 10^{-2}$ )	T/P	0.02	25.62	-0.16	33.80
Citywide demand (difference between 15 and 16 days ago) ( $\times 10^{-2}$ )	T/P	0.05	25.72	-0.15	34.07
Citywide demand (difference between 22 and 23 days ago) ( $\times 10^{-2}$ )	T/P	0.05	25.86	-0.17	34.34
Citywide demand (difference between 29 and 30 days ago) ( $\times 10^{-2}$ )	T/P	0.07	26.00	-0.18	34.62

<sup>a</sup>We use type N for neighboring information, P for periodical information, and T for trend information

2020). In the used models, we incorporate four auxiliary features, (i) temporal features, (ii) meteorological features, (iii) built environment features, and (iv) sociodemographic features. Temporal and meteorological

features vary across different days (dynamic data); built environment and sociodemographic features are static for each census tract, in addition to the road network and

**Table 3** Variables summary statistics aggregated by census tracts

Variable	Unit	Austin		Louisville	
		Mean	StD	Mean	StD
<i>Demographics<sup>a</sup></i>					
Population	Count	2372.16	1692.12	2023.63	821.10
Median income	10 <sup>3</sup> \$	65.24	30.18	44.49	25.56
High education	(%)	0.47	0.20	0.27	0.21
White ethnicity	(%)	0.51	0.23	0.58	0.33
Male	(%)	0.52	0.06	0.48	0.04
Age under 18	(%)	0.17	0.08	0.21	0.08
Age between 18 and 29	(%)	0.26	0.16	0.2	0.11
Age between 30 and 39	(%)	0.20	0.06	0.14	0.05
Age between 40 and 59	(%)	0.23	0.07	0.25	0.05
Age over 60	(%)	0.14	0.07	0.20	0.08
<i>Modes used to travel to work</i>					
Drove alone	(%)	0.70	0.09	0.73	0.12
Taxi	(%)	0.01	0.01	0.03	0.04
Transit	(%)	0.05	0.05	0.07	0.07
Walked	(%)	0.03	0.06	0.04	0.05
Work from home	(%)	0.08	0.05	0.04	0.03
Bicycle	(%)	0.02	0.02	0.01	0.01
Carpooled	(%)	0.09	0.06	0.08	0.04
<i>Infrastructure</i>					
Census tract area	km <sup>2</sup>	2.57	2.12	2.10	1.76
Sidewalk length	km	37.01	18.36	4.54	7.95
Number of signal	Number	8.21	10.97	8.12	12.52
Length of bike lane	km	1.14	2.55	2.79	4.89
Bikesharing station	Number	0.25	0.98	0.34	1.92
Distance to downtown <sup>d</sup>	km	7.16	3.77	5.61	2.59
LTAI		5.50	0.76	5.51	0.82
<i>Land use<sup>b</sup></i>					
Residential	(%)	0.43	0.16	0.42	0.18
Civic	(%)	0.08	0.12	0.10	0.09
Commercial	(%)	0.07	0.07	0.08	0.07
Industrial	(%)	0.03	0.07	0.06	0.09
Mixed	(%)	0.01	0.02	0.00	0.00
Office	(%)	0.04	0.05	0.00	0.00
Parks	(%)	0.08	0.09	0.07	0.13
<i>Meteorological</i>					
Max temperature	°F	70.44	16.17	61.18	19.05
Mini temperature	°F	70.44	16.17	61.18	19.05
Precipitation <sup>c</sup>		0.02	0.13	0.03	0.18
Wind speed	km/h	5.70	3.81	6.93	4.39

<sup>a</sup>For the definition of the different demographics, please refer to the United States Census Bureau ([ensus.org](https://www.census.gov))

<sup>b</sup>For the definition of the different land uses categories, please refer to the American Planning Association APA ([planning.org](https://www.planning.org))

<sup>c</sup>Rainy days coded as (1), and non-rainy days coded as (0)

<sup>d</sup>Measured from the centroid of each tract

infrastructure attributes. Table 3 shows the summary of the statistics of the used auxiliary features.

### Base Models

This subsection introduces four ML techniques that we applied using the proposed methodological framework. We choose the models based on four different types of ML. Linear regression (LR) depends on the assumptions of the linear relationship between the features and the outputs. Support vector regression (SVR) uses a kernel method to impose the non-linearity of the data; gradient-boosting decision tree models the data using an ensemble of if-else rule sets based on tree representation. Finally, we used a deep learning technique to capture the non-linearity of the relationship between the features and the output. We explain the details of each of these models as follows;

**Linear Regression (LR)** LR is a classical machine learning model that assumes a linear or affine relationship between input features and output labels. The simple linear regression takes the following formulation,

$$f(x_i) = w'x_i + b, \tag{1}$$

where  $w \in \mathbb{R}^m$  is the coefficient vector, and  $b \in \mathbb{R}$  is the intercept. The residual  $y_i - f(x_i)$  is assumed to follow a Gaussian distribution. Also assuming the independence of training samples, the parameters can be estimated through the least squares method, which is equivalent to maximum likelihood estimation. It aims at minimizing the sum of squared error, formulated as follows,

$$\min_{w,b} \sum_i (y_i - f(x_i))^2. \tag{2}$$

**Support Vector Regression (SVR)** SVR is an extension of ordinary support vector machine (SVM) for solving regression problems, which was originally designed for classification. To make binary classification, SVM adopts a separating hyperplane  $w'x + b = 0$  to split the feature space  $\mathcal{X}$  into two half-spaces. In the regression case, the hyperplane is turned into a real-valued function  $f(x_i) = w'x_i + b$  resembling to linear regression. Instead of least squares, SVR is trained based on the  $\epsilon$ -insensitive loss, as formulated below,

$$\ell_\epsilon(z_i) = \begin{cases} 0, & \text{if } |z_i| \leq \epsilon, \\ |z_i| - \epsilon, & \text{otherwise,} \end{cases} \tag{3}$$

where  $z_i = y_i - f(x_i)$ . Unlike squared loss in least squares, there is no penalty when the absolute prediction error is not greater than threshold  $\epsilon$ . The complete optimization objective of SVR is given by,



$$\min_{w,b} C \sum_i \ell_\epsilon(z_i) + \frac{\|w\|^2}{2}, \quad (4)$$

where  $C > 0$  is a trade-off coefficient between the  $\epsilon$ -insensitive loss and the regularization term (Scholkopf and Smola 2001). In addition to the linear formulation, SVR can deal with non-linearity by introducing the kernel trick, which projects the input features into a high-dimensional space using a kernel function  $\phi(\cdot)$ ; thus the prediction function becomes  $f(x) = w' \phi(x) + b$ . Common kernel functions include the polynomial kernel, spline kernel, and Gaussian kernel (Scholkopf and Smola 2001; Wendland 2004). Despite the strong expressive power of non-linear kernels, the time complexity of model estimation can be worse than  $O(N^3)$  (Platt 1998).

**Gradient boosting decision tree (GBDT)** Decision tree (DT) has a superior prediction performance and good interpretability (Wu et al. 2008). Each decision rule corresponds to an exclusive path from the root node to a leaf node in the tree, while each leaf node is associated with a group of samples in the training set. The rule set of a DT actually partitions a subspace  $\mathcal{S}$  of the feature space  $\mathcal{X}$  into many sub-regions. For each input feature vector, DT searches for the sub-region to which this vector belongs, and prediction can be made based on the samples associated with the leaf node in the corresponding decision rule. The training process of a DT is a search for a satisfactory set of decision rule, i.e., a partition of  $\mathcal{S}$ . It has early been proven that finding an optimal rule set for a DT is *NP-Complete* (Kwiatkowski et al. 1992); hence a greedy heuristic algorithm is often used for model training and the resulting DT is suboptimal. But, concerning a DT for regression problems with a determined feature space partition, the optimal output value of a specific leaf node can be concluded as the average labels of all the associated samples (Bishop 2006). Therefore, it is noteworthy to mention that the output of a DT is limited between the minimum and maximum of labels of its training data. Despite the inability in extrapolation, DT is competitive over many other machine learning models in interpolation. Despite the boom of deep learning research in the recent decade, it has been found that deep learning is not a panacea for all tasks. For example, Shwartz-Ziv and Armon (2022) pointed out that DT outperforms deep models on many tabular data, which is the case for the auxiliary information in our experiment. In addition, according to a research (Bojer and Meldgaard 2021) on the winning solutions of forecasting competitions on Kaggle,<sup>7</sup> DT is the most competitive

one over other machine learning models. Similar results can be observed from the latest survey conducted by Kaggle<sup>8</sup>; DT is the most popular method among its users on top of linear models. To further improve its generalization ability and reduce the risk of over-fitting, ensemble learning is combined with DT, and GBDT is one of the representatives (Friedman 2001). The principal idea of Boosting is to express the model as a summation of multiple base models. There are a number of improvements made on GBDT in terms of engineering implementation, including XGBoost (Chen and Guestrin 2016), CatBoost (Dorogush et al. 2017) and LightGBM (Ke et al. 2017). In this paper, we adopt LightGBM, a highly efficient GBDT framework, which utilizes two specially designed techniques, namely Gradient-based One-Side Sampling and Exclusive Feature Bundling, to ease the computational burden of large-scale data involved in model training without sacrificing the prediction accuracy.

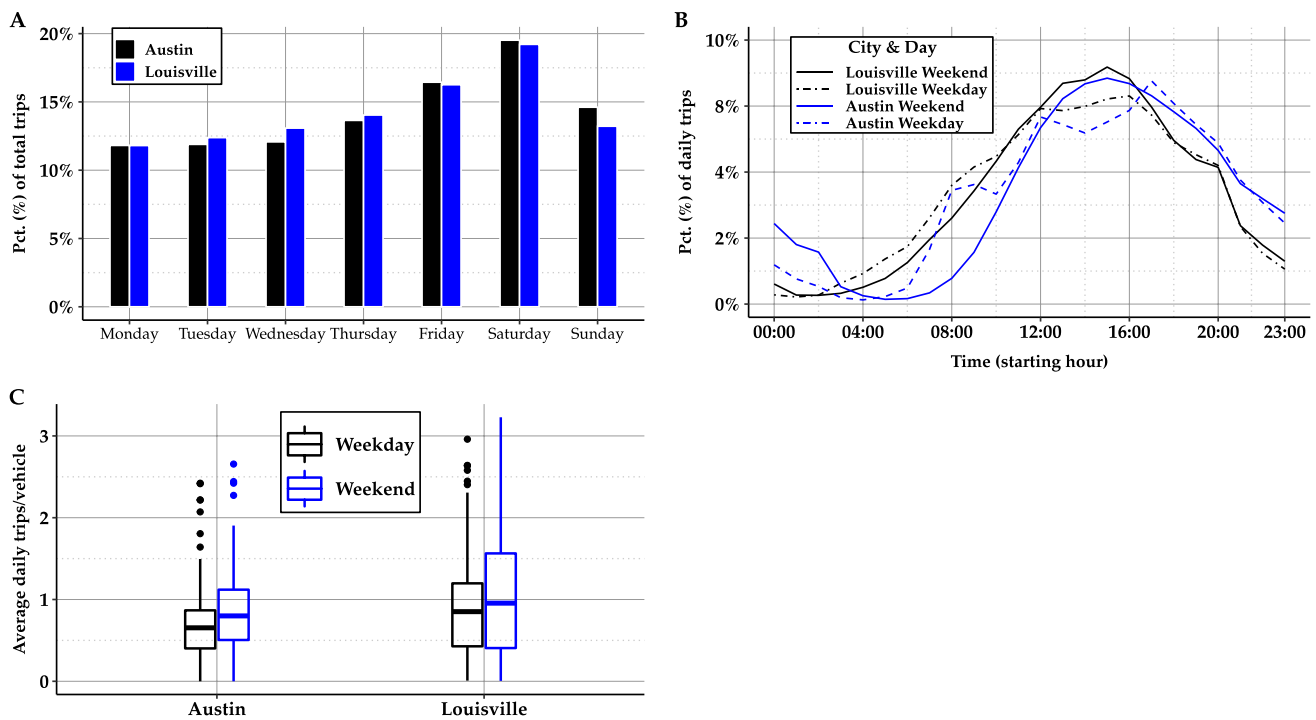
**Long Short-term Memory Neural Network (LSTM NN)** LSTM is a recurrent neural network (RNN) model for modelling sequential data. In contrast to most non-recurrent neural networks, RNN allows loop connections in its architecture, which feed the outputs of a layer to itself as its inputs in the following time step (LeCun et al. 2015). An ordinary RNN layer maintains a hidden state  $\mathbf{H}$  along time; in each time step  $t$ , it is fed with the current feature vector  $\mathbf{x}_t$  and the previous hidden state  $\mathbf{H}_{t-1}$ . The hidden state of time step  $t$  is updated by the non-linear transformations of the two inputs, while the output is given by another non-linear transformation of the hidden state. LSTM improves RNN's ability of modelling long-term relationship by introducing three gated units (i.e., input gate, output gate, and forget gate) and an additional memory state  $\mathbf{C}$  in the recurrent layer. The three gated units apply different non-linear transformations on the two inputs, whereby the memory state and the hidden state are also updated,

$$\begin{aligned} I_t &= \phi(w'_I x_t + v'_I h_{t-1} + b_I), \\ O_t &= \phi(w'_O x_t + v'_O h_{t-1} + b_O), \\ F_t &= \phi(w'_F x_t + v'_F h_{t-1} + b_F), \\ C_t &= F_t \otimes C_{t-1} + I_t \otimes \psi(w'_C x_t + v'_C h_{t-1} + b_C), \\ H_t &= O_t \otimes \phi(C_t). \end{aligned} \quad (5)$$

where  $\phi(\cdot)$  and  $\psi(\cdot)$  are sigmoid and hyperbolic tangent activation functions, respectively;  $w$ ,  $v$  and  $b$  are parameters;  $\otimes$  is the Hadamard product. The training of LSTM NN can be realized via back-propagation through time, which unfolds

<sup>7</sup> A platforms for hosting data science competitions (kaggle.com).

<sup>8</sup> <https://www.kaggle.com/competitions/kaggle-survey-2022>, accessed 25/11/2022.



**Fig. 2** Average **A** daily, **B** hourly demand distribution, **C** and fleet utilization (number of trips per vehicle) daily distribution. Weekend includes Saturday and Sunday, weekdays are the rest of the days

the computation steps along time to allow the use of the chain rule (Hochreiter and Schmidhuber 1997).

**Model Transfer**

Time series differencing is used to remove trends from the data (detrend) in response to GBDT’s defect in extrapolation. Denote two consecutive time series values as  $x_{i-1}$  and  $x_i$ , the first-order differencing yields a transformed label  $y_i$  as follows,

$$y_i = x_i - x_{i-1}. \tag{6}$$

However, differencing alone is inadequate regarding the model transfer problem due to the uneven distributions between data in  $D_S$  and  $D_T$ . Figure 2 compares the distributions of average daily trips per vehicle between two cities. The demand pattern of Louisville has a more significant dispersion than Austin’s. If Austin was used as the source city and Louisville as the target city, the trained GBDT model might underestimate the demand in Louisville, as the model would not learn much information about high daily trip demand, and this was the main reason to use Austin, as the source city; which is called covariate shift. Covariate shift refers to the case when the probability distributions between

training data and test data differ while the conditional distributions of labels on input data are the same (Sugiyama and Kawanabe 2012). Nevertheless, an implicit assumption of standard supervised learning models, including GBDT, is that the training and test data follow the same probability (Ben-David et al. 2010), refraining from dealing with covariate shift.

$\mathcal{D}_S$  and  $\mathcal{D}_T$  denote the distributions of data in  $D_S$  and  $D_T$ , respectively; denote the actual underlying functions mapping input feature vectors to labels on the two sets of data as  $f_S$  and  $f_T$ . Then, following Ben-David et al. (2010), we call  $\langle \mathcal{D}_S, f_S \rangle$  the *source domain* and  $\langle \mathcal{D}_T, f_T \rangle$  the *target domain*. The expected error on the source domain can be obtained by

$$e_S(g, f_S) = \mathbb{E}_{x \sim \mathcal{D}_S} [\ell(g(x), f_S(x))], \tag{7}$$

where  $g(\cdot)$  is the model,  $\ell(\cdot, \cdot)$  is the loss function. Similarly, the expected error on the target domain can be defined as  $e_T(g, f_T)$ .

In general, models are trained to minimize the empirical error on the source domain; nevertheless, in the model transfer problem, we minimize the error on the target domain. One option is transforming the data from  $x$  to  $x'$  such that the corresponding distributions  $\mathcal{D}_S$  is similar to  $\mathcal{D}_T$ . Inspired

by the batch normalization strategy in deep learning (Ioffe and Szegedy 2015), we proposed the sample normalization strategy to transport the knowledge learned from the source time series to the target time series. We implicitly assumed that time-series dynamics irrespective of the value scale conditional on given features. For each sample, before extracting time series features, the input time series segment was normalized to a mean of zero and a variance of one. Denote the time series segment as  $\mathbf{t}$ , the normalized segment  $\tilde{\mathbf{t}}$  can be obtained by:

$$\tilde{\mathbf{t}} = \frac{\mathbf{t} - E(\mathbf{t})}{D(\mathbf{t})}, \quad (8)$$

where  $E(\mathbf{t})$  and  $D(\mathbf{t})$  are the mean and the standard deviation of  $\mathbf{t}$  respectively.

Sample normalization was adopted to reduce the covariate shift for the studied model transfer problem. The feature construction procedure with sample normalization was presented in Algorithm 1. Feature vectors are constructed for data in each census tract following the FEATURECONSTRUCTION procedure. It should be noted that a complete training sample consists of a feature vector and a label, where the label also needs normalization. Recall that sample normalization takes a time series segment of consecutive ( $w$ ) days; the label corresponds to the day right after this segment and needs to be normalized using the mean and standard deviation of the previous segment. As the label represents the day when the demand is predicted, it should not be combined with the previous segment when calculating the normalization parameters, i.e., the mean and standard deviation.<sup>9</sup>

---

### Algorithm 1 Feature Construction (FEATURECONSTRUCTION) with Sample Normalization

---

```

1: procedure SAMPLENORM( $\mathbf{t}$ )
2:   ▷  $\mathbf{t}$  is a time series segment.
3:    $\mu \leftarrow \frac{1}{|\mathbf{t}|} \sum_{i=1}^{|\mathbf{t}|} t_i$ 
4:    $\sigma \leftarrow \left( \frac{1}{|\mathbf{t}|} \sum_{i=1}^{|\mathbf{t}|} (t_i - \mu)^2 \right)^{1/2}$ 
5:    $\tilde{\mathbf{t}} \leftarrow (\mathbf{t} - \mu) / \sigma$ 
6:   return  $\tilde{\mathbf{t}}$ 
7: end procedure

1: procedure FEATURECONSTRUCTION( $w, \mathbf{t}, \mathbf{x}_{tm}, \mathbf{x}_{bs}$ )
2:   ▷  $w$  is the time window size9,  $\mathbf{t}$  is a time series.
3:   ▷  $\mathbf{x}_{tm}$  is the temporal and meteorological information.
4:   ▷  $\mathbf{x}_{bs}$  is the built environment and sociodemographical information.
5:   Initiate  $\mathbf{s} \leftarrow \emptyset$ 
6:   for  $i \leftarrow 1$  to  $|\mathbf{t}| - w$  do
7:      $\tilde{\mathbf{t}} \leftarrow \text{SAMPLENORM}(\mathbf{t}^{(i:i+w)})$ 
8:      $\tilde{\mathbf{r}} \leftarrow \text{FEATUREEXTRACTION}(\tilde{\mathbf{t}})$ 
9:      $\mathbf{u} \leftarrow \tilde{\mathbf{r}} \cup \mathbf{x}_{bs} \cup \mathbf{x}_{tm}^{(i+w)}$ 
10:     $\mathbf{s} \leftarrow \mathbf{s} \cup \{\mathbf{u}\}$ 
11:   end for
12:   return  $\mathbf{s}$ 
13: end procedure

```

▷  $i$  is the index of day.  
 ▷ Normalize the time series segment.  
 ▷ Extract time series features from  $\tilde{\mathbf{t}}$ .  
 ▷  $\mathbf{u}$  is the feature vector of a sample.

---

## Data Collection and Processing

We predicted the scooter's fleet utilization using different data sources; we used fleet utilization daily rates (daily number of trips per vehicle) of one city to predict the fleet utilization in the other city. The primary datasets are the

scooter trip booking data from Austin; Texas, and Louisville; Kentucky; in Austin, the data spanned from April 2018 to January 2020, while in Louisville, the collected data spanned from August 2018 to January 2020. We used trip data in combination with other open-source data, specifically; (i) census sociodemographics information obtained

<sup>9</sup> The window size used for feature extraction is 28 days. In this paper, we assume there are only 1 month available in the target city, hence the choice of window size approximately 1 month.

**Table 4** Summary of trip characteristics per city

	Mean	StD	Min	$Q^{1st}$	Median	$Q^{3rd}$	Max	City
Distance (km)	1.57	1.50	0.10	0.64	1.13	1.96	45.71	Austin
Distance (km)	2.07	2.26	0.10	0.64	1.29	2.61	32.19	Louisville
Duration (min)	11.09	11.83	1.00	4.45	7.20	12.92	120.00	Austin
Duration (min)	15.62	17.22	1.00	5.00	9.00	19.00	120.00	Louisville
Speed (km/h)	9.76	4.59	0.06	6.39	9.29	12.68	25.00	Austin
Speed (km/h)	9.05	4.47	0.07	5.79	8.59	12.00	25.00	Louisville
Fleet utilization (trip/vehicle)	0.71	0.40	0.00	0.44	0.69	0.93	2.66	Austin
Fleet utilization (trip/vehicle)	0.91	0.61	0.00	0.42	0.86	1.26	3.23	Louisville

Number of trips after the data cleaning process is in Austin = 7,038,490 trip, and in Louisville = 389,739 trip

from the United States Census Bureau<sup>10</sup> census data include sociodemographic information aggregated per census tract including age group, gender ratio, race and ethnicity, marital status, education level, household income, house price, and modes used to work. (ii) Built environment and infrastructure data from ([openstreetmap.org](https://openstreetmap.org)); this data set included all the physical features of the built environment and POI, such as roads, bike lanes, intersections (both signalized and non-signalized), bikesharing stations, shops, banks, and educational institutes. (iii) Meteorological data from ([visua.lcrossing.com](https://visua.lcrossing.com)), and this dataset included hourly atmospheric temperature, wind speed, rain, and snow conditions.

The collected data was processed and cleaned in several steps. Austin's data set contained scooter and e-bike trips; the latter were removed. Other procedures were similar to what McKenzie (2019), Liu et al. (2019), Zou et al. (2020) used were implemented to remove false entries and false trips; all trips shorter than 100 m and longer than 50 km were removed; trips shorter than one minute and more extended than 2 h duration were removed. Also, we removed trips with speeds higher than 25 mph. Build environment, infrastructure, and POI data were aggregated per census tract as it is the spatial aggregation unit. It is to be noted that we did not observe any problems in the sociodemographic census information and meteorological data.

## Case Study

Scooter booking data from Louisville, Kentucky, and Austin, Texas, were obtained from the open city portals (Austin Shared Mobility Services 2022; Louisville Open Data 2022). Scooter fleet size in both cities is different, wherein Austin, the maximum fleet size is 15,000 vehicles (Austin Shared Mobility Services 2022) and in Louisville, it is 1200 vehicles (Louisville Open Data 2022). Also, the operational regulations for shared-e-scooters are slightly different in

both cities. The speed limit is 20 mph in Austin, while it is 15 mph in Louisville; helmet use is advised and mandatory for under-18 users in Austin, while it is mandatory for all users in Louisville.

It is essential to mention that the two cities in this study are different in terms of population, where Austin's population (0.98 million) is approximately 1.5 times the population of Louisville. Also, Austin has several options for public transportation, compared to Louisville, which has only bus service; however, they have similarities in terms of modal share for work trips, as both cities are car-dependent cities, wherein Louisville (89%) of work trips are done in private cars, compared to 81.2 % in Austin ([census.gov](https://census.gov)).

## Analysis Results

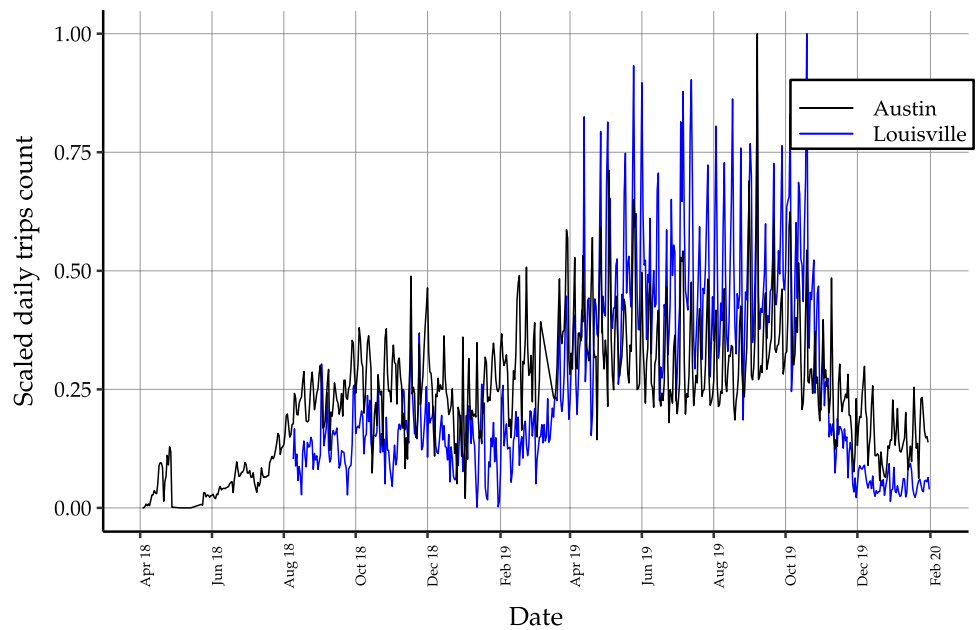
As discussed in the literature review, "Literature Review" section, some of the challenges faced by the shared mobility service are directly linked to the spatial and temporal demand pattern; therefore, we start by analyzing the trip characteristics and then the demand patterns temporally and spatially to recognize the patterns in both cities, and compare them to define similarities and differences, after that, we show the results of the estimated models and its adequacy for fitting the data.

### Trips Characteristics

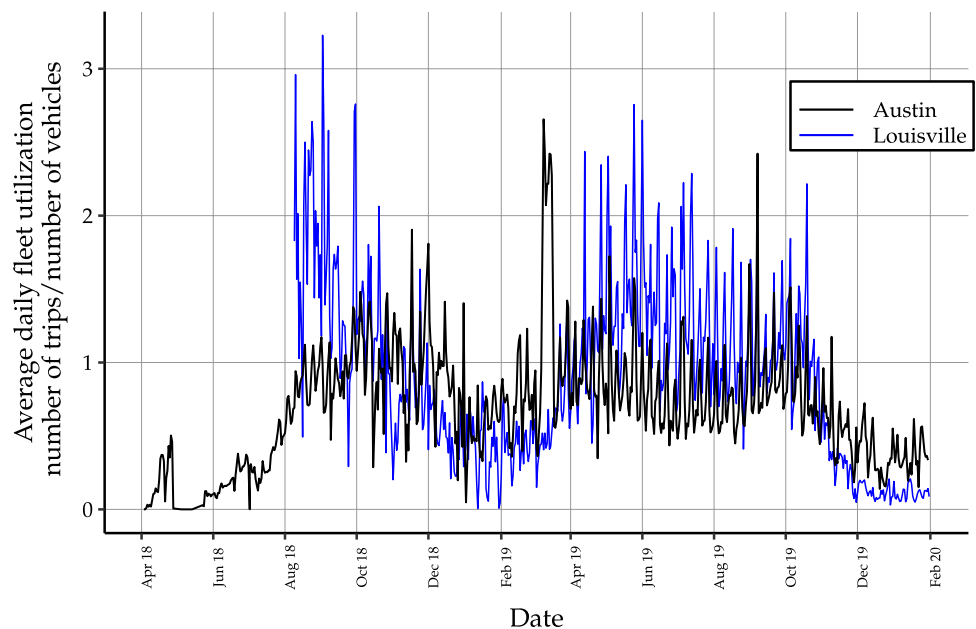
After cleaning the data as discussed in "Methods, Data, and Case Study" section, the original 9 million trips in Austin were reduced to around 7 million trips (78% of original trips), and the initial 500,000 in Louisville were reduced to approximately 390,000 trips (77% of original trips). We analyzed and compared the characteristics of the cleaned trips in both cities as shown in Table 4. To investigate the differences and similarities between trips characteristics in both cities, a  $t$  test of the mean of two samples is performed for the trip distance, duration, and speed between the two

<sup>10</sup> :census.gov, accessed 5 March 2022.

**Fig. 3** Daily demand distribution, scaled demand



**Fig. 4** Utilized fleet daily distribution, number of trips per vehicle per day



cities, and it shows that the difference between the two samples is significant for the three trip metrics ( $P < 0.0001$ ). Louisville trips tend to be longer in distance and duration, but Austin trips tend to be faster. We investigated the average street slopes in both cities using Google Earth Engine (Gorelick et al. 2017) to explore if slopes impact the mean scooter trips' speed, with no significant difference found as both cities have almost flat terrain except for some localized areas. We also removed the demand data for the second week of March 2019 from the Austin dataset, as the SXSW music festival took place at that time, and the demand was (5–6)

times the average demand for the rest of the data collection period, as extreme outlier removal is essential for improving model performance (Saum et al. 2020).

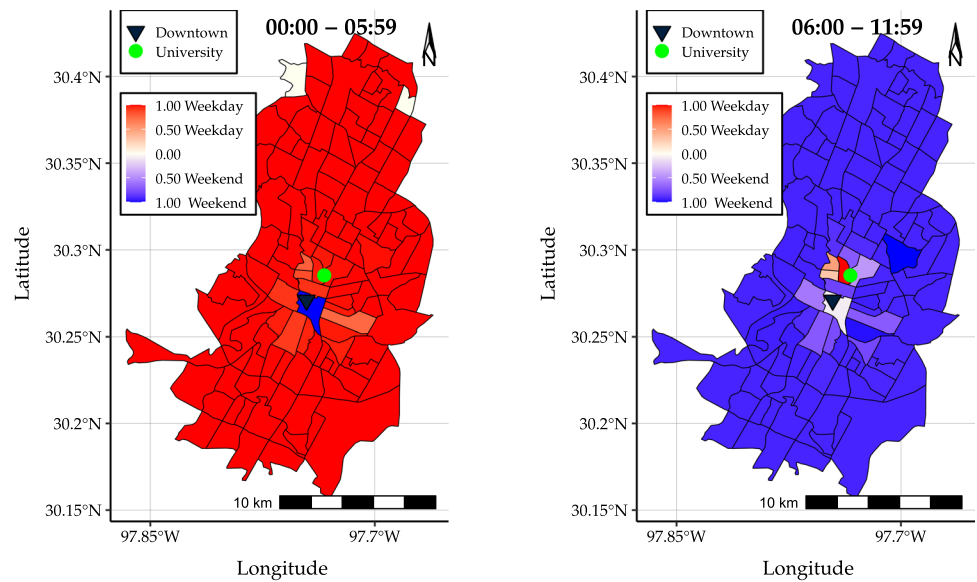
## Demand Analysis

### Temporal Analysis

To compare the demand trends and patterns in the two cities, we normalized the demand by scaling the daily demand; we divided it by the maximum number of trips for the

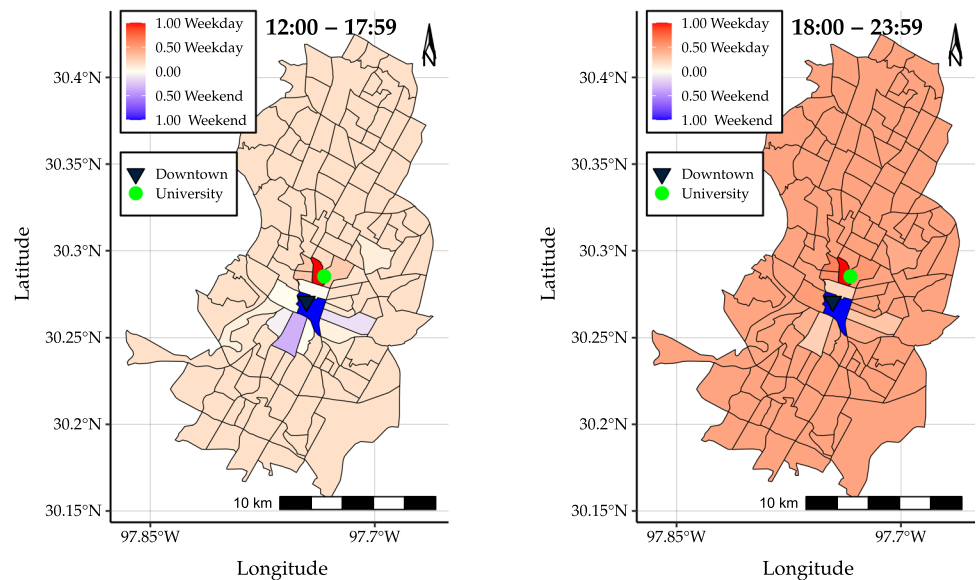


**Fig. 5** Spatial distribution of the dominant difference between weekends and weekdays trips, Austin, TX



(a) Early morning hours: 00:00 - 05:59

(b) Morning hours: 06:00 - 11:59



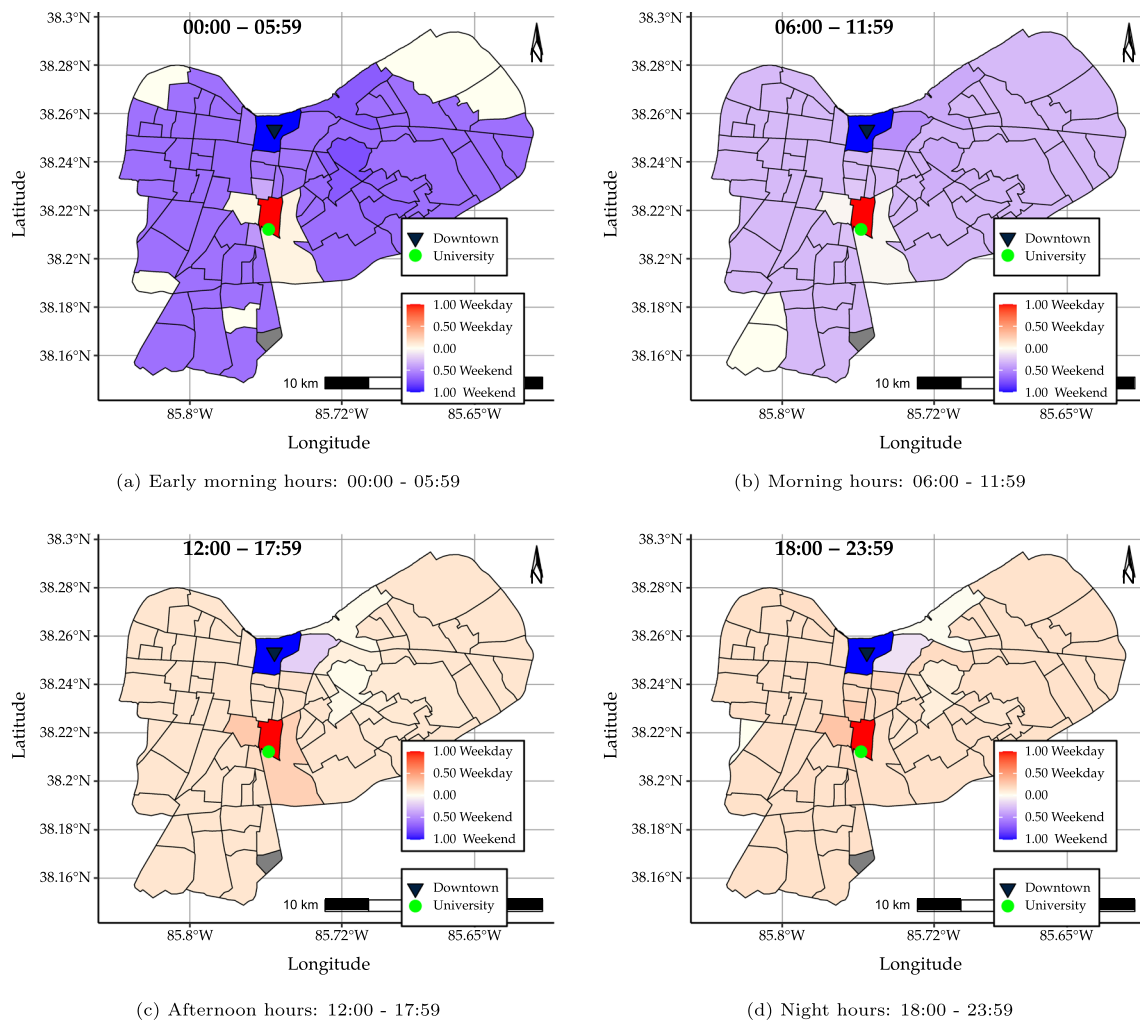
(c) Afternoon hours: 12:00 - 17:59

(d) Night hours: 18:00 - 23:59

investigated duration for each city, as shown in Fig. 3; similar to the procedures used by Schmöller et al. (2015). The scaled demand shows similar trends in both cities, where the demand increases in spring and summer, and it starts to drop from October (autumn) and continues to decline until January, when the lowest average of the year is observed. Following the same scaling procedures to control the different fleet sizes, we also compared the number of trips per vehicle in both cities. Interestingly, the number of trips per vehicle fluctuates in a different trend than the demand, with Louisville having higher trips per vehicle than Austin at the beginning of the deployment period, i.e., the first two months, and

it decreases for the following 6 months. It almost matches in both cities for almost 7 months in 2019 (from April till November), despite the different fleet sizes in both cities. It is necessary to mention that the fleet size was not fixed during the data collection period, as mentioned and considered from Louisville Open Data (2022); Austin Shared Mobility Services (2022). Figure 4 shows the average daily number of trips per vehicle in both cities.

We also investigated the hourly and daily demand. In general, the demand patterns for weekdays are similar in both cities, where the demand per weekday as a percentage of the total weekly demand is stable from Mondays



**Fig. 6** Spatial distribution of the dominant difference between weekends and weekdays trips, Louisville, KY

to Thursdays, with a slight increase on Fridays. The peak of the demand happens on Saturdays (around 20% of the total weekly demand and 50% more than the average weekday demand). Sundays’ demand is slightly higher than the weekdays’ demand, excluding Fridays; refer to Fig. 2. This trend in demand distribution shows an increase in scooter usage during weekends, which primarily indicates the use of scooters for leisure-related activities, which is the principal purpose of scooters (Abouelela et al. 2021a).

The last temporal element to investigate was the hourly demand. We aggregated the hourly demand for weekdays and weekends. We found that the hourly demand in both cities follows similar trends. In Austin, the weekday demand is a left-skewed Bimodal distribution, with one minor peak around 8:00 and the other peak of the day between 12:00 and 17:00, which is the primary demand peak. The morning peak (around 8:00) does not exist on weekends, and the only demand peak is around 13:00. The morning peak hour during the weekdays could indicate that scooters are used

for commuting trips at this time of the day. It is also to be noticed that there is a high demand for trips on the weekends’ early morning hours, which might indicate the use of leisure trips at these times of the day. In Louisville, the trends are similar except that there is no morning peak hour demand during weekdays, and there is a low number of early morning trips during the weekends; refer to Fig. 2.

**Spatial Analysis**

We analyzed the spatial demand in the two cities at different periods of the day and on different weekdays guided by the temporal analysis results. The demand was aggregated per each census tract, per weekday v.s. weekend, per time of the day. We divided the day into four primary time intervals, each 6 h long, as shown in Figs. 5 and 6. We investigated the weekday demand dominance by normalizing the difference between the number of weekday trips and weekend trips and scaled the difference from (-1 to +1). First,

**Table 5** Models performance metrics

Label Differencing	Sample Normalization	Performance ( $\times 10^{-5}$ )				
		Model	Train RMSE	Train MAE	Test RMSE	Test MAE
<i>Models without transfer learning</i>						
–	–	LightGBM	531.6	82.9	2195.7	<b>382.6</b>
–	–	LR	1017.2	185.5	2164.1	388.3
–	–	SVR	1064.8	242.9	<b>2092.5</b>	440.4
–	–	LSTM	1333.8	360.2	2366.2	484.9
<i>Models after transfer learning</i>						
✓	✗	LightGBM	469.3	97.9	2291.9	394.1
✗	✓	LightGBM	1059.0	174.1	2037.6	390.5
✓	✓	LightGBM	873.9	130.7	<b>1845.6</b>	<b>346.8</b>
✓	✗	LR	1017.3	185.7	2168.5	389.6
✗	✓	LR	1166.8	178.4	<b>2034.7</b>	<b>378.7</b>
✓	✓	LR	1263.6	185.4	2054.4	381.2
✓	✗	SVR	1064.0	215.1	2135.6	449.5
✗	✓	SVR	1212.7	181.4	2200.8	381.4
✓	✓	SVR	1296.4	177.8	<b>2208.3</b>	<b>371.3</b>
✓	✗	LSTM	1274.3	284.6	2647.4	515.8
✗	✓	LSTM	1176.5	182.1	2677.9	480.6
✓	✓	LSTM	1140.6	179.3	<b>2376.0</b>	<b>436.4</b>

Bold value indicates the best-performing models

we assigned each trip to the starting census tract; then we aggregated the trips temporally to the weekend trips that happened on Saturday and Sunday and weekday trips for the other five days of the week; then we calculated the difference between the average weekend, and weekday trips per week for each tract ( $i = \text{number of weeks} \dots 1 - m$ ,  $j = \text{number of tracts} \dots 1 - n$ ), after this we normalized the difference by dividing by the maximum difference for each tract;

$$T_{ij} = X \cdot \sum_{i=1, j=1}^{i=m, j=n} \frac{\text{Avg}(\text{Weekend}_{ij}) - \text{Avg}(\text{Weekday}_{ij})}{\max(\text{Avg}(\text{Weekend}_{ij}) - \text{Avg}(\text{Weekday}_{ij}))}, \tag{9}$$

where  $i = \text{number of weeks}$ ,  $j = \text{number of tract}$ , and<sup>11</sup>

$$X = \begin{cases} -1, & \text{if } \text{Avg}(\text{Weekend}_{ij}) - \text{Avg}(\text{Weekend}_{ij}) \leq 0, \\ 1, & \text{otherwise.} \end{cases} \tag{10}$$

The following spatio-temporal trends are observed. In Austin, the downtown area is mainly dominated by weekend trips at different times except before noon hours, where there is almost no difference in demand between weekday and weekend trips. A similar trend is noticed in the University of Texas at Austin area, which is dominated by weekday trips at different times of the day. Weekend trips dominate only the

early morning hours, while the rest of the day is dominated by weekday trips by different ratios.

The spatio-temporal analysis shows interesting findings; both cities' downtown and university areas are two major attraction areas, and their spatial and temporal demand patterns are the same regardless of their use in the rest of the city. In Louisville, similar trends were also noticed. Weekend trips dominate the downtown area, and the University of Louisville area is dominated by weekday trips at all times of the day. For the rest of the city, the early morning hours

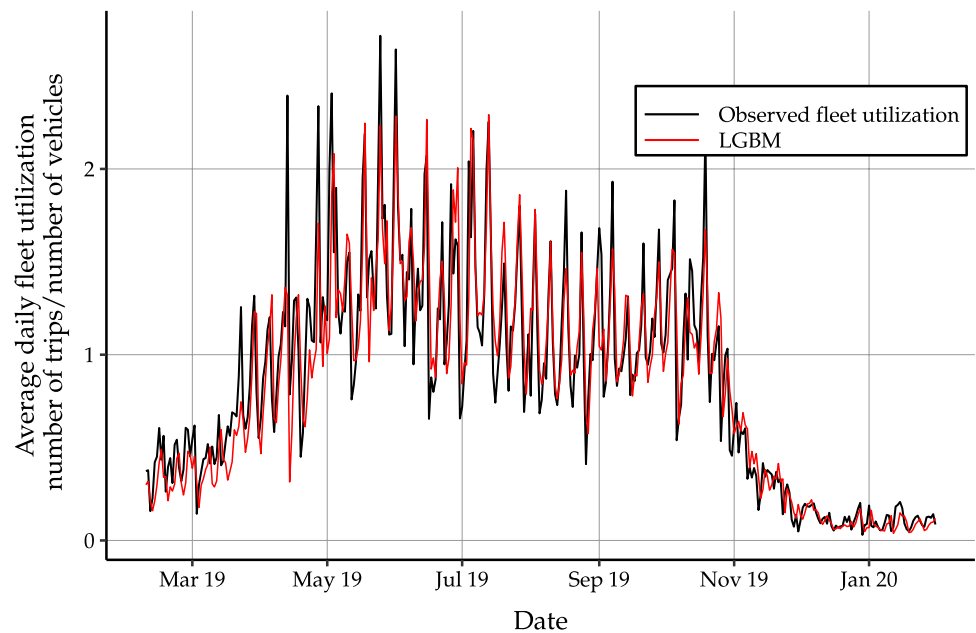
**Table 6** Relative feature importance of top-10 features, LightGBM

Rank	Feature	Relative importance (%)
1	Demand (previous day)	6.6
2	Elapsed days since operation	6.3
3	Temperature	5.3
4	Day of the week	3.5
5	Demand (average of the past week)	3.1
6	Demand (difference between previous day and 2 days ago)	3.1
7	Demand (range of the past week)	3.1
8	Demand (7 days ago)	2.5
9	Distance to downtown	2.1
10	Citywide demand (difference between previous day and 2 days ago)	2.1

<sup>11</sup> X here is used for plotting positive values on the scale of Figs. 5 and 6.

**Table 7** Relative feature importance of feature groups, LightGBM

Feature group	Importance (%)	Performance after feature removal ( $\times 10^{-5}$ )			
		Train RMSE	Train MAE	Test RMSE	Test MAE
Time series features	67.0	1243.7	199.8	2637.6	457.3
Temporal features	9.8	978.2	146.6	1976.0	362.4
Sociodemographical features	9.6	887.7	133.9	1873.3	350.4
Meteorological features	7.1	1051.2	169.0	2182.9	399.1
Built environment features	6.6	876.4	133.3	1869.3	349.7

**Fig. 7** Observed versus predicted fleet utilization using LightGBM

till before noon are dominated by weekend trips, and the rest of the day is dominated by weekday trips with different ratios. Although the temporal distribution of trips is almost the same in both cities, their geographical distribution is different, which we believe is due to the different urban structures of both cities.

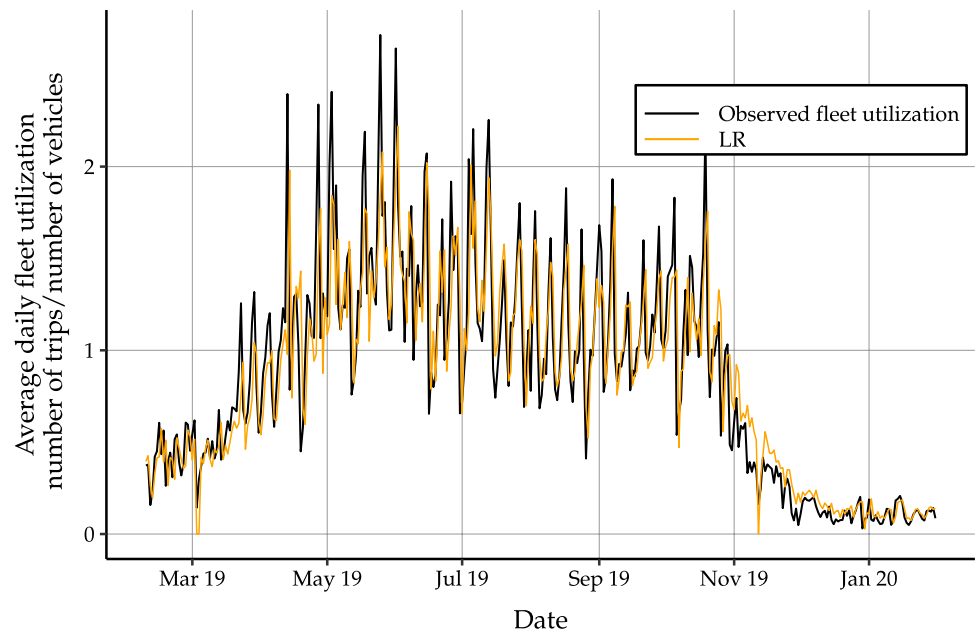
### Model Results, and Performance Evaluation

We investigated the effectiveness of the proposed methodology used for the model transfer problem for demand prediction. The prediction accuracy was evaluated using two metrics; root mean squared error (RMSE), and mean absolute error (MAE). The proposed framework was applied to the different used ML techniques. We first compare the performance of the models as shown in Table 5 upper part, and then we compared the performance of the model after the transfer (label differencing, and sample normalization), Table 5 lower part.

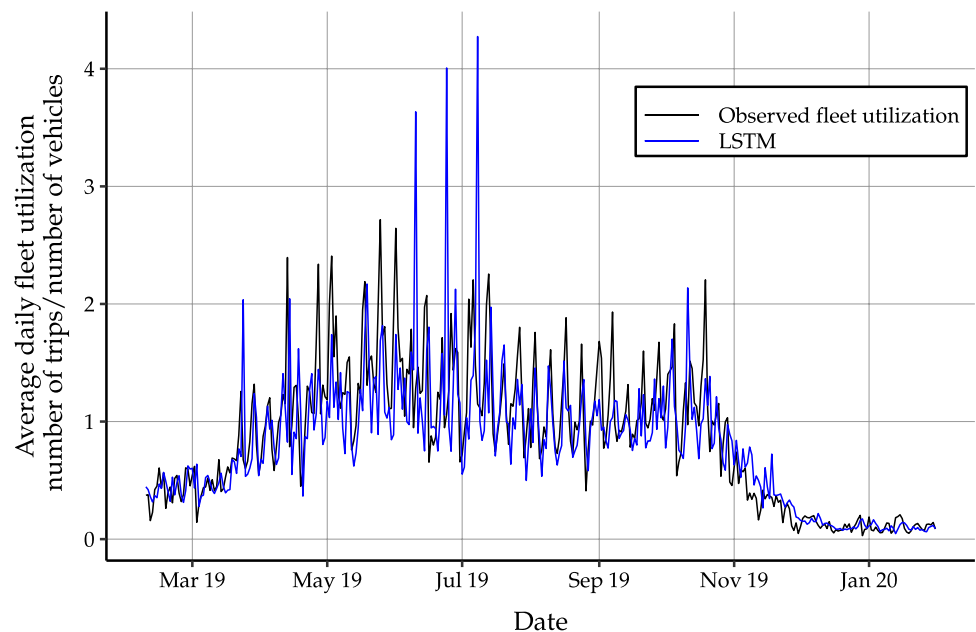
To further improve the transferability of the model, we applied the model transfer strategies to all the models. We

applied the different transfer strategies as shown in Table 5 lower part, which shows the model's prediction results summary after applying the different transfer strategies. A time series prediction without treatment of the covariate shift issue suffers from low RMSE and MAE on the training set. However, when faced with unseen data in another city, the test set's performance suffers considerably because of distinct time series patterns. Firstly, we applied label differencing, but it did not improve accuracy as the distribution inconsistency in the input space was not addressed; similarly, only applying sample normalization was ineffective. The transfer error was finally reduced when the two strategies were used simultaneously, which is evident in the best-performing model, LightGBM. For LightGBM, The RMSE dropped from 2195.7 to 1845.6, which showed an improvement of the performance by approximately 15.9%. Meanwhile, a drop in accuracy on the training set was also observed, indicating a less severe over-fitting model; in other words, the proposed method was satisfactory in improving the generalization ability and robustness of the model in the

**Fig. 8** Observed versus predicted fleet utilization using Linear Regression



**Fig. 9** Observed versus predicted fleet utilization using LSTM NN



transfer learning problem. Further error analysis is presented in the following section.

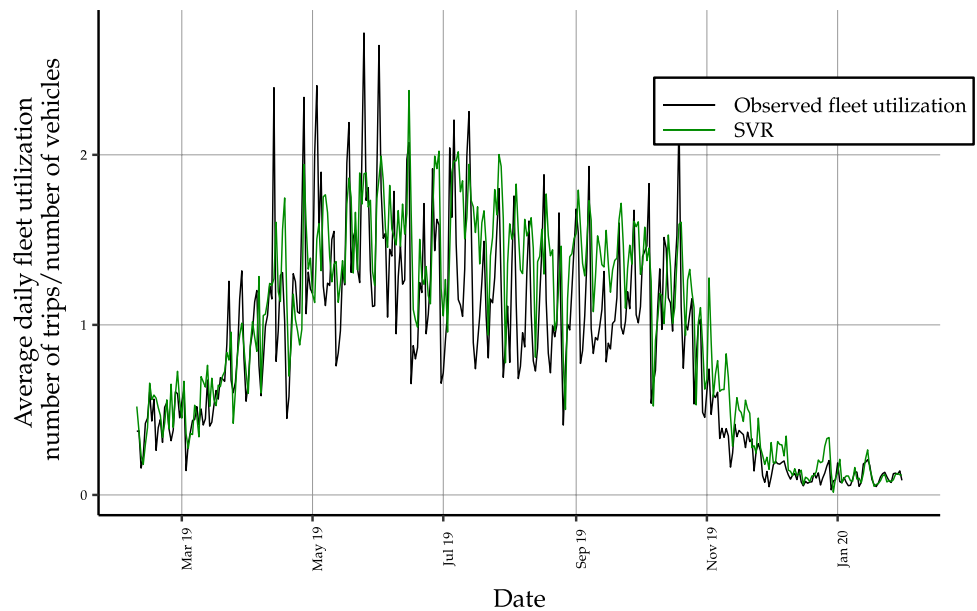
As The LightGBM model was the model with the best prediction performance, we evaluated the importance of factors influencing the prediction using the number of node splits corresponding to each feature in the trained LightGBM model. The more a feature was adopted for a split in the tree, the higher its contribution to the prediction (Liu et al. 2020). We ranked features by their relative importance, the top 10 listed in Table 6. As a time series prediction model, lagged demand values and their statistics are essential to

the prediction, where the one-day lagged demand contributed the most, accounting for 6.6% of all feature splits in the trained decision trees. Among the top 10 features, three time-varying and one census tract-related auxiliary features, i.e., elapsed days since operation,<sup>12</sup> temperature, day of the

<sup>12</sup> The term “elapsed days since operation” here means the number of days from the first operation day of the service to the day corresponding with the sample to be predicted. This feature is used as the demand pattern of a shared mobility service can differ between its starting stage and later.



**Fig. 10** Observed versus predicted fleet utilization using SVR



week, and distance to downtown significantly contributed to the prediction.

To quantify the influence of the different factors groups, we categorized the features into five main groups, refer to Table 7. Time series features accounted for 67.0% of node splits in the trees, whereas each category of auxiliary features accounted for approximately 6–10% of node splits. Further experiments were performed to see whether removing specific feature groups would significantly reduce prediction accuracy. We found that removing every feature group will more or less negatively impact the model performance. The results are generally consistent with their relative importance; the removal of time series features—the most critical group of features—resulted in a performance drop of around 43% in Test RMSE. Removing auxiliary features did not incur severe impacts, where the accuracy reduction caused by removing built environment features or sociodemographic features was less than 2% per group.

### Error Analysis Description

We analyzed the prediction error, its value distribution, temporal distribution, and spatial location considering the test set Louisville's dataset. The actual and predicted number of trips per vehicle per day were plotted along the temporal axis in Figs. 7, 8, 9, and 10.

It can be observed that all the estimated models captured the overall demand pattern with some shortcomings. The LR model tends to overestimate the utilization rate between (1–1.75) vehicles per trip, and it underestimates the demand when it is higher than 1.75 trips per day; for the rest of the value, it is somehow able to estimate the fleet utilization rate.

SVR was consistently unable to predict the utilization rate; for rates below 1.25 vehicle/trip, the model underestimated the results, and for rates over 1.245 vehicle/trip, the model overestimated the utilization rates. Regarding the temporal distribution of the error, Fig. 10, SVR was the model with the least prediction capabilities. LSTM could not accurately predict the low utilization rate and tended to overestimate the utilization below 1.2 trips per day and underestimated the demand higher than 1.2; also, the model had some incidents where the estimated utilization rates were significantly higher than the actual rate. The LGBM model had the best performance among the four models. It can be observed that the prediction results of the proposed model capture most of the demand seasonal peaks and troughs dynamics without lag except for the several sudden spikes in the early stage of operation (e.g., the spike in mid-April). However, the model inclines underestimation regarding peak values, possibly an outcome of model regularization, as predictions of large values are more likely to be connected with high errors (error terms are increasingly proportional to the absolute demand value). Potential solutions include increased training data and additional information like special events and fine-grained weather forecasts. To be able to observe and understand the previous prediction trends, we plotted the predicted values in comparison to the actual values using three different graphs, where Fig. 11 shows the distribution of the predicted values in reference to the actual observed utilization rate, Fig. 12 shows the distribution of the difference between the observed utilization rate and the predicted utilization rate. Finally, Fig. 13 shows the predicted

values against the actual values and the corresponding regression line.

We also spatially analyzed the difference between the predicted and actual fleet utilization rate; we plotted the difference between the average utilization rate and the predicted one per census tract. Figure 14 shows the prediction error spatial distribution; it can be observed that for all models, except SVR, the errors in most census tracts are low or even zero. SVR overestimates the utilization rate in all the tracts except the university area, where it underestimates the utilization rate. The errors have a distinctive pattern for the other three models, mostly occurring around downtown and the university area's tracts. The high demand can explain the relatively high errors in these two areas, and error is proportional to the absolute demand value and more events that make predictions difficult.

The conclusion of the error analysis process, which was done in multiple dimensions, shows that the LGBM model is superior in prediction accuracy compared to the other used ML models, including LSTM.

## Discussion, and Conclusion

### Discussion

The observed spatio-temporal scooters' demand patterns in the two examined cities show the demand and fleet utilization rate seasonality. The demand is different for the different hours of the day and the day of the week; also, the spatial demand distribution is different in the two examined cities, and it depends on the time of the day. Nevertheless, there is a significant spatial common phenomenon in both cities, with the demand spatially concentrated around the downtown and the university areas. Therefore, system operations such as vehicles' supply management and allocation and redistribution should consider such patterns in the deployment and redistribution process and ensure deploying the number of vehicles in the desired locations that are changing according to the actual demand. The used framework shows a simplified and effective way to predict the number of trips per vehicle (fleet utilization) for one of the rapidly expanding shared mobility services, shared-e-scooter, depending on open-source data. This framework could be used (after testing) for similar dockless, free-floating micromobility shared systems, which exhibited similar travel behavior, e.g., free-floating bike-sharing services (Zhu et al. 2020; McKenzie 2019). Moreover, similar data characteristics to the one used in this study should be available for other shared mobility services to implement the used framework; for each trip, trip starting and ending spatial points, starting and ending timings, trip speed, and trip distance. As the methodology section explains, the framework depends on employing the historical

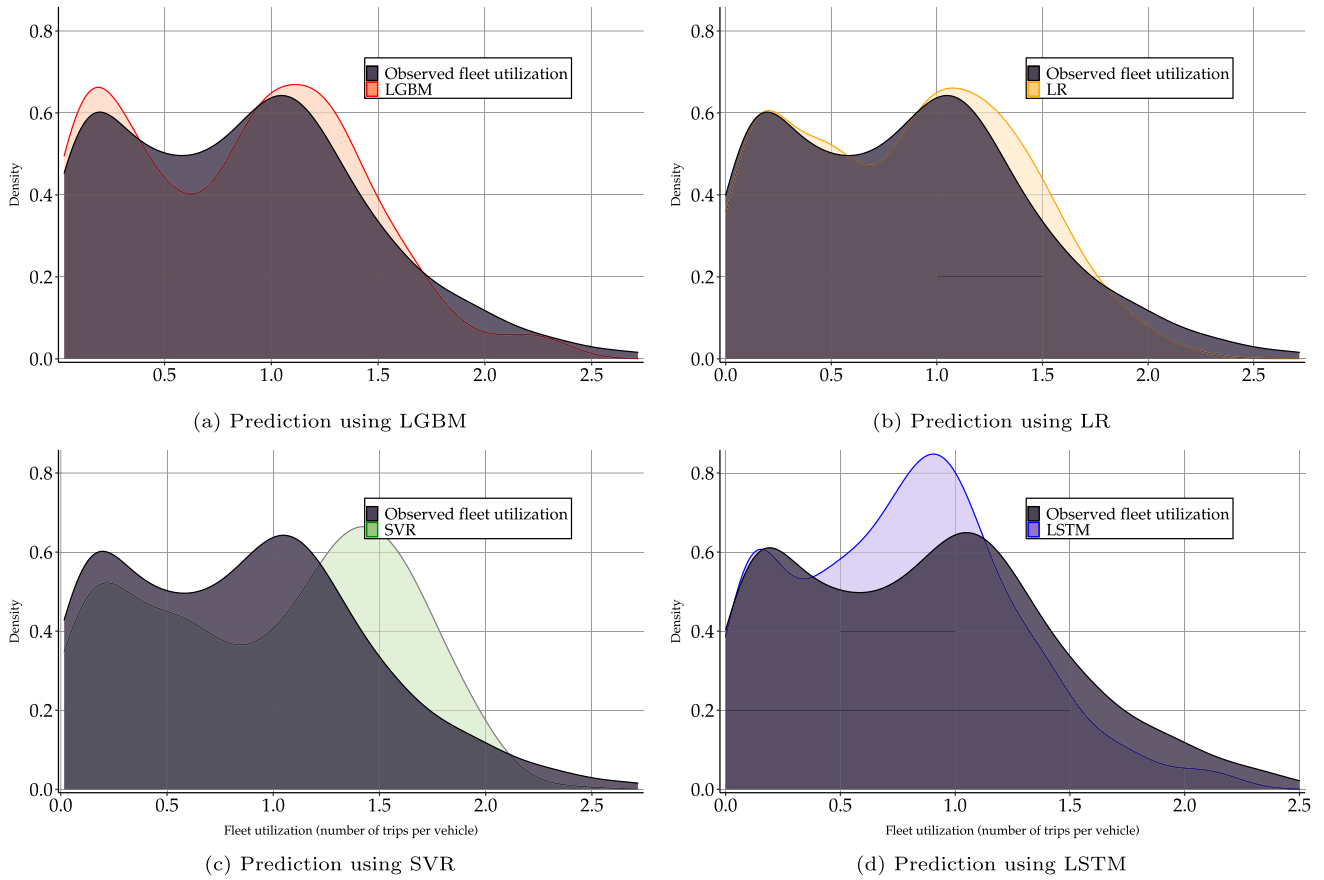
demand data combined with open-source data; therefore, different stakeholders could use the framework to predict the daily number of trips per vehicle and deploy the vehicles in the expected locations accordingly. The error analysis section ("Methods, Data, and Case Study" section) shows that the increase in the number of days used in the prediction process increases the accuracy of the models; therefore, the continuous use of such models would improve the model accuracy over time. It is also to be noticed that we used the ridership (the number of trips per vehicle per day) for the prediction task for two main reasons; firstly, we wanted to control the fleet size in both cities to be able to compare the demand and to normalize the impact of the supply. Secondly, demand is directly tied to supply in the case of shared mobility services, and estimating absolute demand will lead to a biased estimation (Gammelli et al. 2020). Moreover, the predicted fleet utilization rates should decide the fleet size. Table 4 shows that the median and mean daily average number of trips per vehicle in the two cities are under one trip/vehicle/day; therefore, more investigating measures need to be applied to define the reasons behind the low ridership. In addition, cities should study the consequences of making ridership rates a compelling factor for the number of deployed scooters. Based on our analysis, we believe fleet size should be dynamically decided, if not daily, which needs further research to determine its efficiency in vehicle balancing and redistribution and the generated additional VKT weekly according to the seasons. Special event periods, such as the SXSW music festival in Austin, should consider different supply and vehicle rebalancing operation schemes due to the increased demand compared to regular condition days.

### Conclusion

The methodology and data show a promising approach that the stakeholders could implement and use to organize scooters and similar shared micromobility vehicle services. However, the model needs to be tested for the other service to validate user behavior differences. Also, publishing the trip booking data publicly by cities should be encouraged as it plays a vital role in encouraging researchers from industry and academia to investigate such services use behavior and discover innovative methods to enhance service operations.

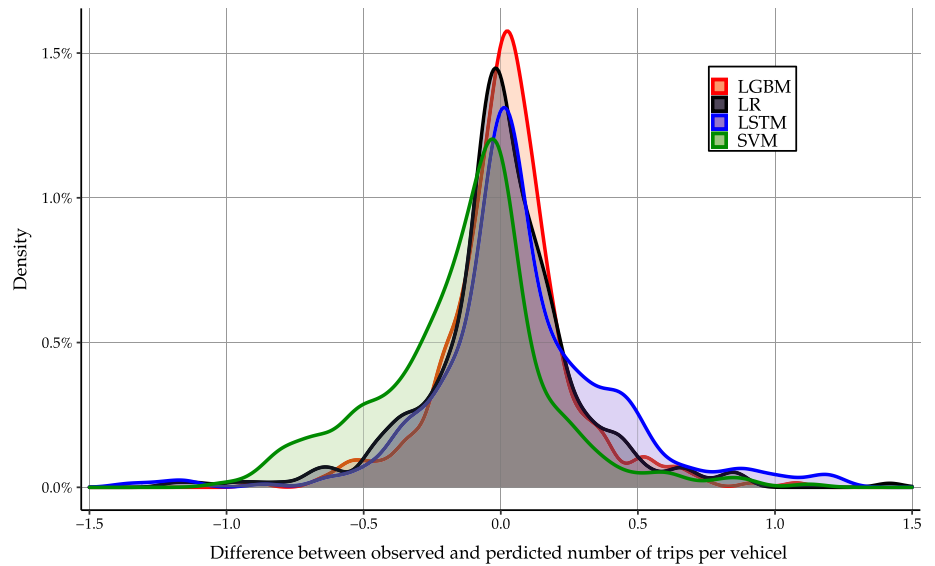
## Appendix 1: Additional Analysis

See Figs. 11, 12, 13 and 14.



**Fig. 11** Observed fleet utilization distribution versus predicted fleet utilization using different ML models

**Fig. 12** Distribution of the difference between (Actual - predicted average daily trips per vehicle)



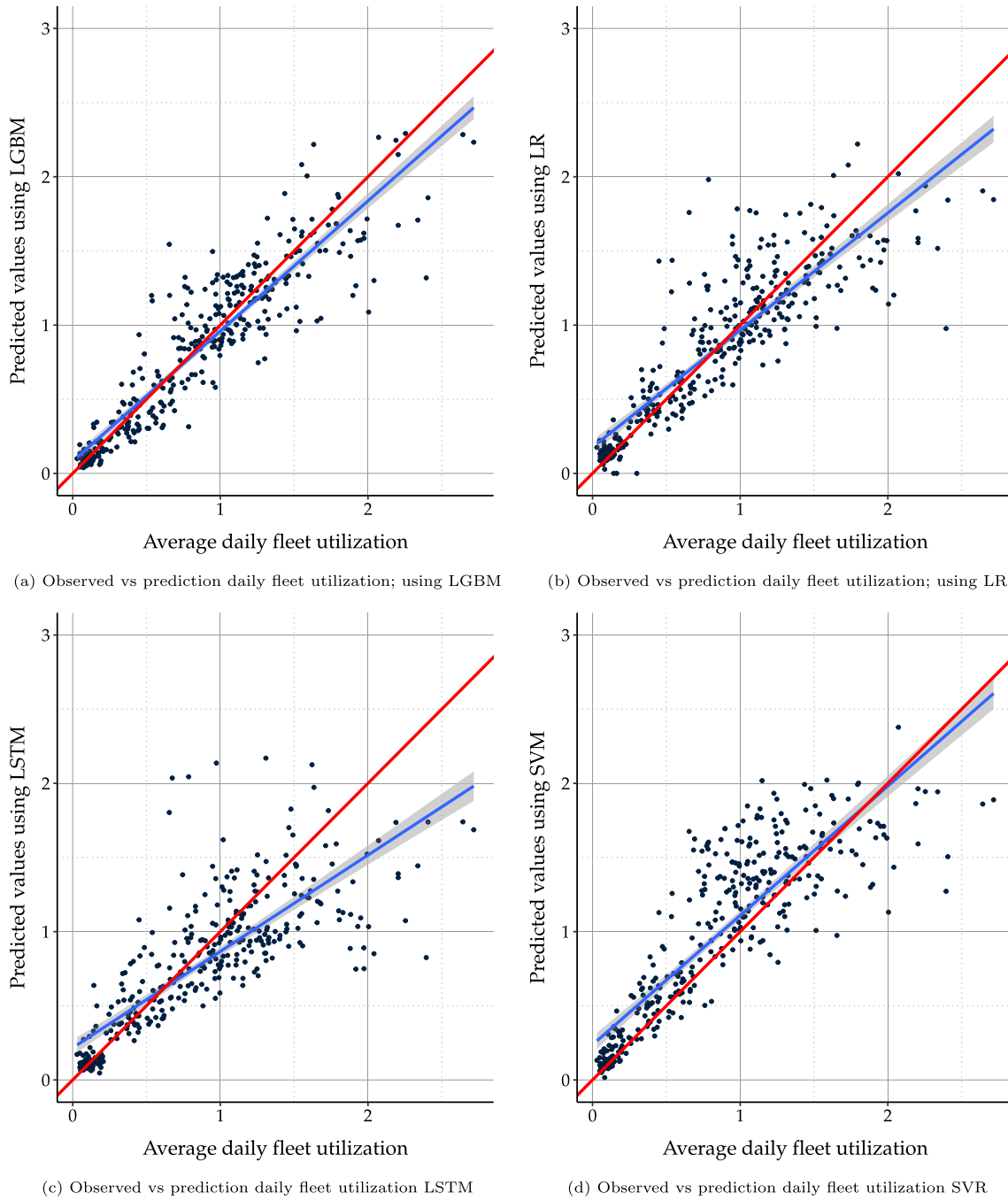
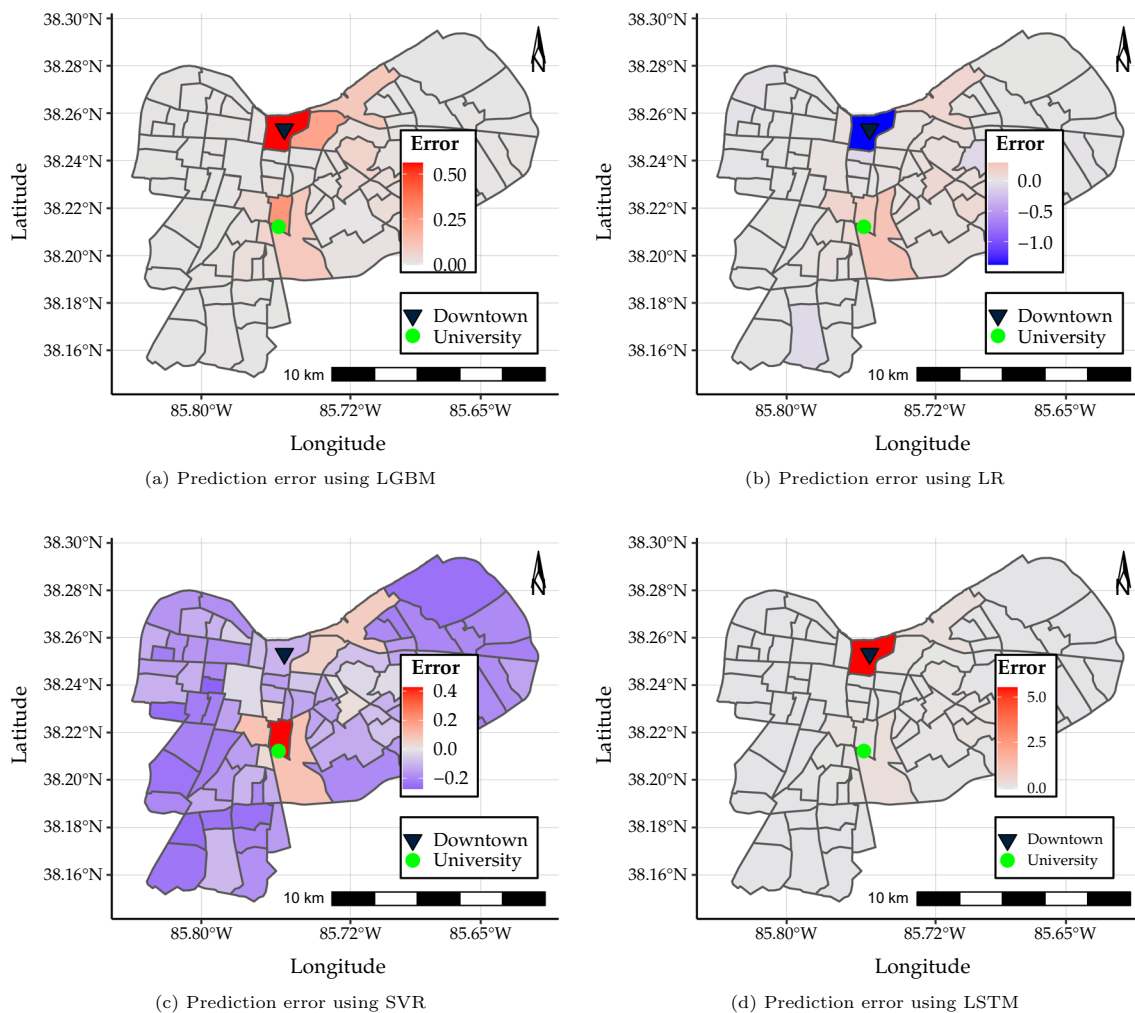


Fig. 13 Observed average daily fleet utilization versus predicted fleet utilization, the blue line is regression line and the red line is 1:1 slope line



**Fig. 14** Spatial distribution of the mean difference between observed fleet utilization and predicted fleet utilization

**Acknowledgements** This study was partially funded by the DAAD Project Number 57474280 Verkehr-SuTra: Technologies for Sustainable Transportation, within the Programme: A New Passage to India – Deutsch-Indische Hochschulkooperationen ab 2019, the German Federal Ministry of Education and Research, Bundesministerium für Bildung und Forschung (BMBF), project FuturTrans: Indo-German Collaborative Research Center on Intelligent Transportation Systems.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated

otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abouelela M, Al Haddad C, Antoniou C (2021a) Are e-scooters parked near bus stops? Findings from Louisville, Kentucky. Findings, p 29001
- Abouelela M, Al Haddad C, Antoniou C (2021b) Are young users willing to shift from carsharing to scooter-sharing? *Transp Res Part D Transp Environ* 95:102821
- Abouelela M, Tirachini A, Chaniotakis E, Antoniou C (2022) Characterizing the adoption and frequency of use of a pooled rides service. *Transp Res Part C Emerg Technol* 138:103632
- Abouelela M, Chaniotakis E, Antoniou C (2023) Understanding the landscape of shared-e-scooters in North America; spatiotemporal



- analysis and policy insights. *Transp Res Part A Policy Pract* 169:103602
- Austin Shared Mobility Services (2022) <http://austintexas.gov/departments/shared-mobility-services>. Accessed 3 Mar 22
- Baek K, Lee H, Chung J-H, Kim J (2021) Electric scooter sharing: how do people value it as a last-mile transportation mode? *Transp Res Part D Transp Environ* 90:102642
- Becker H, Balac M, Ciari F, Axhausen KW (2020) Assessing the welfare impacts of shared mobility and mobility as a service (MaaS). *Transp Res Part A Policy Pract* 131:228–243
- Ben-David S, Blitzer J, Crammer K, Kulesza A, Pereira F, Vaughan JW (2010) A theory of learning from different domains. *Mach Learn* 79:151–175. <https://doi.org/10.1007/s10994-009-5152-4>
- Bhattacharya A, Romani M, Stern N (2012) Infrastructure for development: meeting the challenge. CCCEP, Grantham Research Institute on Climate Change and the Environment and G, 24
- Bishop CM (2006) Pattern recognition and machine learning. In: Information science and statistics. Springer, New York
- Bojer CS, Meldgaard JP (2021) Kaggle forecasting competitions: an overlooked learning opportunity. *Int J Forecast* 37:587–603
- Cantelmo G, Kucharski R, Antoniou C (2020) Low-dimensional model for bike-sharing demand forecasting that explicitly accounts for weather data. *Transp Res Rec* 2674:132–144
- Cerutti PS, Martins RD, Macke J, Sarate JAR (2019) Green, but not as green as that: an analysis of a Brazilian bike-sharing system. *J Clean Prod* 217:185–193
- Chaniotakis E, Efthymiou D, Antoniou C (2020) Data aspects of the evaluation of demand for emerging transportation systems. In: Demand for emerging transportation systems. Elsevier, pp 77–99
- Chaniotakis E, Abouelela M, Antoniou C, Goulias K (2022) Investigating social media spatiotemporal transferability for transport. *Commun Transp Res* 2:100081
- Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM Press, San Francisco, pp 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chen Y-W, Cheng C-Y, Li S-F, Yu C-H (2018) Location optimization for multiple types of charging stations for electric scooters. *Appl Soft Comput* 67:519–528
- Choi J, Yoon J (2017) Utilizing spatial big data platform in evaluating correlations between rental housing car sharing and public transportation. *Spat Inf Res* 25:555–564
- Circella G, Alemi F, Tiedeman K, Handy S, Mokhtarian PL et al (2018) The adoption of shared mobility in California and its relationship with other components of travel behavior. Technical Report National Center for Sustainable Transportation
- De Lorimier A, El-Geneidy AM (2013) Understanding the factors affecting vehicle usage and availability in carsharing networks: a case study of communauto carsharing system from Montréal, Canada. *Int J Sustain Transp* 7:35–51
- Degele J, Gorr A, Haas K, Kormann D, Krauss S, Lipinski P, Tenbih M, Koppenhoefer C, Fauser J, Hertweck D (2018) Identifying e-scooter sharing customer segments using clustering. In: 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp 1–8. <https://doi.org/10.1109/ICE.2018.8436288>
- Dorogush AV, Ershov V, Gulin A (2017) CatBoost: gradient boosting with categorical features support. In: Workshop on ML systems at the 31st conference on neural information processing systems. Curran Associates Inc., Long Beach, pp 1–7
- Durán-Rodas D, Chaniotakis E, Wulfhorst G, Antoniou C (2020a) Open source data-driven method to identify most influencing spatiotemporal factors. An example of station-based bike sharing. In: Mapping the travel behavior genome. Elsevier, pp 503–526
- Durán-Rodas D, Villeneuve D, Pereira FC, Wulfhorst G (2020b) How fair is the allocation of bike-sharing infrastructure? Framework for a qualitative and quantitative spatial fairness assessment. *Transp Res Part A Policy Pract* 140:299–319
- El-Assi W, Mahmoud MS, Habib KN (2017) Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. *Transportation* 44:589–613
- Estache A (2010) Infrastructure finance in developing countries: an overview. *EIB Pap* 15:60–88
- Fearnley N, Johnsson E, Berge SH (2020) Patterns of e-scooter use in combination with public transport. Findings
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 29:1189–1232
- Gammelli D, Peled I, Rodrigues F, Pacino D, Kurtaran HA, Pereira FC (2020) Estimating latent demand of shared mobility through censored Gaussian processes. arXiv preprint [arXiv:2001.07402](https://arxiv.org/abs/2001.07402)
- Gao X, Lee GM (2019) Moment-based rental prediction for bicycle-sharing transportation systems using a hybrid genetic algorithm and machine learning. *Comput Ind Eng* 128:60–69
- Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R (2017) Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens Environ*. <https://doi.org/10.1016/j.rse.2017.06.031>
- Haghighat AK, Ravichandra-Mouli V, Chakraborty P, Esfandiari Y, Arabi S, Sharma A (2020) Applications of deep learning in intelligent transportation systems. *J Big Data Anal Transp* 2:115–145
- Heineke K, Kloss B, Scurtu D, Weig F (2019) Sizing the micro mobility market. <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/micromobilitys-15000-mile-check-up>. Accessed 7 Mar 21
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9:1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hu S, Chen P, Lin H, Xie C, Chen X (2018) Promoting carsharing attractiveness and efficiency: an exploratory analysis. *Transp Res Part D Transp Environ* 65:229–243
- Iliashenko O, Iliashenko V, Lukyanchenko E (2021) Big data in transport modelling and planning. *Transp Res Procedia* 54:900–908
- Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on Machine Learning ICML'15. JMLR, Lille, pp 448–456. <https://doi.org/10.5555/3045118.3045167>
- Janssen C, Barbour W, Hafkenschiel E, Abkowitz M, Philip C, Work DB (2020) City-to-city and temporal assessment of peer city scooter policy. *Transp Res Rec* 2674:219–232
- Jiang Z, Mondschein A (2021) Analyzing parking sentiment and its relationship to parking supply and the built environment using online reviews. *J Big Data Anal Transp* 3:61–79
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y (2017) LightGBM: a highly efficient gradient boosting decision tree. In: Proceedings of the 31st international conference on neural information processing systems. Curran Associates, Inc., Long Beach, pp 3146–3154
- Kim D, Shin H, Im H, Park J (2012) Factors influencing travel behaviors in bikesharing. In: Transportation Research Board 91st annual meeting
- Kim D, Ko J, Park Y (2015) Factors affecting electric vehicle sharing program participants' attitudes about car ownership and program participation. *Transp Res Part D Transp Environ* 36:96–106
- Ko J, Ki H, Lee S (2019) Factors affecting carsharing program participants' car ownership changes. *Transp Lett* 11:208–218
- Kostic B, Loft MP, Rodrigues F, Borysov SS (2021) Deep survival modelling for shared mobility. *Transp Res Part C Emerg Technol* 128:103213

- Kwiatkowski D, Phillips PC, Schmidt P, Shin Y (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. *J Econom* 54:159–178. [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y)
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444. <https://doi.org/10.1038/nature14539>
- Lin P, Weng J, Liang Q, Alivanistos D, Ma S (2018) Impact of weather conditions and built environment on public bikesharing trips in Beijing. In: *Networks and spatial economics*, pp 1–17
- Liu D, Dong H, Li T, Corcoran J, Ji S (2018) Vehicle scheduling approach and its practice to optimise public bicycle redistribution in Hangzhou. *IET Intell Transp Syst* 12:976–985
- Liu M, Seeder S, Li H et al (2019) Analysis of e-scooter trips and their temporal usage patterns. *Inst Transp Eng ITE J* 89:44–49
- Liu Y, Lyu C, Liu X, Liu Z (2020) Automatic feature engineering for bus passenger flow prediction based on modular convolutional neural network. *IEEE Trans Intell Transp Syst* 22:2349–2358. <https://doi.org/10.1109/TITS.2020.3004254>
- Liu X, Van Hentenryck P, Zhao X (2021a) Optimization models for estimating transit network origin-destination flows with big transit data. *J Big Data Anal Transp* 3:247–262
- Liu Y, Lyu C, Zhang Y, Liu Z, Yu W, Qu X (2021b) DeepTSP: deep traffic state prediction model based on large-scale empirical data. *Commun Transp Res* 1:100012
- Louisville Open Data (2022) <https://data.louisvilleky.gov/dataset/dockless-vehicles>. Accessed 24 Jan 22
- Luo M, Wen H, Luo Y, Du B, Klemmer K, Zhu H (2019) Dynamic demand prediction for expanding electric vehicle sharing systems: a graph sequence learning approach. arXiv preprint [arXiv:1903.04051](https://arxiv.org/abs/1903.04051)
- Luo H, Zhang Z, Gkritza K, Cai H (2021) Are shared electric scooters competing with buses? A case study in Indianapolis. *Transp Res Part D Transp Environ* 97:102877
- Lyu C, Wu X, Liu Y, Liu Z, Yang X (2020) Exploring multi-scale spatial relationship between built environment and public bicycle ridership: a case study in Nanjing. *J Transp Land Use* 13:447–467. <https://doi.org/10.5198/jtlu.2020.1568>
- Mattson J, Godavarthy R (2017) Bike share in Fargo, North Dakota: keys to success and factors affecting ridership. *Sustain Cities Soc* 34:174–182
- McKenzie G (2019) Spatiotemporal comparative analysis of scooter-share and bike-share usage patterns in Washington, DC. *J Transp Geogr* 78:19–28
- Møller T, Simlett J (2020) Micromobility: moving cities into a sustainable future. Technical Report EY
- Müller J, Correia GHdA., Bogenberger K (2017) An explanatory model approach for the spatial distribution of free-floating carsharing bookings: a case-study of German cities. *Sustainability* 9:1290
- NACTO (2019) Shared micromobility in the US:2019. Technical Report National Association of City Transportation Officials. <https://nacto.org/wp-content/uploads/2020/08/2020bikesharesnapshot.pdf>
- Namiri NK, Lui H, Tangney T, Allen IE, Cohen AJ, Breyer BN (2020) Electric scooter injuries and hospital admissions in the United States, 2014–2018. *JAMA Surg* 155:357–359
- Nikitas A, Wallgren P, Rexfelt O (2015) The paradox of public acceptance of bike sharing in Gothenburg. In: *Proceedings of the Institution of Civil Engineers-Engineering Sustainability*, vol 169. Thomas Telford Ltd., pp 101–113
- Platt J (1998) Sequential minimal optimization: a fast algorithm for training support vector machines. Technical Report MSR-TR-98-14 Microsoft Research
- Raux C, Zoubir A, Geyik M (2017) Who are bike sharing schemes members and do they travel differently? The case of Lyon's Velo'v scheme. *Transp Res Part A Policy Pract* 106:350–363
- Reck DJ, Haitao H, Guidon S, Axhausen KW (2021) Explaining shared micromobility usage, competition and mode choice by modelling empirical data from Zurich, Switzerland. *Transp Res Part C Emerg Technol* 124:102947
- Ricci M (2015) Bike sharing: a review of evidence on impacts and processes of implementation and operation. *Res Transp Bus Manag* 15:28–38
- Santacreu A, Yannis G, de Saint Leon O, Crist P (2020) Safe micromobility. Technical Report International Transportation Forum
- Saum N, Sugiura S, Piantanakulchai M (2020) Short-term demand and volatility prediction of shared micro-mobility: a case study of e-scooter in Thammasat University. In: *2020 Forum on Integrated and Sustainable Transportation Systems (FISTS)*. IEEE, pp 27–32
- Schaeffers T (2013) Exploring carsharing usage motives: a hierarchical means-end chain analysis. *Transp Res Part A Policy Pract* 47:69–77. <https://doi.org/10.1016/j.tra.2012.10.024>
- Schmöller S, Weikl S, Müller J, Bogenberger K (2015) Empirical analysis of free-floating carsharing usage: the Munich and Berlin case. *Transp Res Part C Emerg Technol* 56:34–51
- Scholkopf B, Smola AJ (2001) *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT Press, Cambridge
- Sevtsuk A, Basu R, Li X, Kalvo R (2021) A big data approach to understanding pedestrian route choice preferences: evidence from San Francisco. *Travel Behav Soci* 25:41–51
- Shaheen S, Cohen A, Zohdy I et al (2016) *Shared mobility: current practices and guiding principles*. Technical Report United States. Federal Highway Administration
- Shaheen SA, Cohen AP (2013) Carsharing and personal vehicle services: worldwide market developments and emerging trends. *Int J Sustain Transp* 7:5–34
- Shared and Digital Mobility Committee (2018) Taxonomy and definitions for terms related to shared mobility and enabling technologies. Technical Report SAE International. [https://doi.org/10.4271/J3163\\_201809](https://doi.org/10.4271/J3163_201809)
- Shen Y, Zhang X, Zhao J (2018) Understanding the usage of dockless bike sharing in Singapore. *Int J Sustain Transp* 12:686–700
- Shwartz-Ziv R, Armon A (2022) Tabular data: deep learning is not all you need. *Inf Fusion* 81:84–90. <https://doi.org/10.1016/j.inffus.2021.11.011>
- Sperling D (2018) *Three revolutions: steering automated, shared, and electric vehicles to a better future*. Island Press
- Spinney J, Lin W-I (2018) Are you being shared? Mobility, data and social relations in Shanghai's public bike sharing 2.0 sector. *Appl Mobilities* 3:66–83
- Stojanović N, Stojanović D (2020) Big mobility data analytics for traffic monitoring and control. *Facta Univ Ser Autom Control Robot* 19:087–102
- Sugiyama M, Kawanabe M (2012) Machine learning in non-stationary environments: introduction to covariate shift adaptation. In: *Adaptive computation and machine learning*. MIT Press, Cambridge
- Sun Y, Mobasheri A, Hu X, Wang W (2017) Investigating impacts of environmental factors on the cycling behavior of bicycle-sharing users. *Sustainability* 9:1060
- Ting KH, Lee LS, Pickl S, Seow H-V (2021) Shared mobility problems: a systematic review on types, variants, characteristics, and solution approaches. *Appl Sci* 11:7996
- Tirachini A (2020) Ride-hailing, travel behaviour and sustainable mobility: an international review. *Transportation* 47:2011–2047
- Tirachini A, del Río M (2019) Ride-hailing in Santiago de Chile: Users' characterisation and effects on travel behaviour. *Transp Policy* 82:46–57

- Tirachini A, Gomez-Lobo A (2020) Does ride-hailing increase or decrease vehicle kilometers traveled (VKT)? A simulation approach for Santiago de Chile. *Int J Sustain Transp* 14:187–204
- Torre-Bastida AI, Del Ser J, Laña I, Ilardia M, Bilbao MN, Campos-Cordobés S (2018) Big data for transportation and mobility: recent advances, trends and challenges. *IET Intell Transp Syst* 12:742–755
- Turoń K, Czech P, Tóth J (2019) Safety and security aspects in shared mobility systems. *Sci J Silesian Univ Technol Ser Transp* 104:169–175
- United Nations Department of Economic and Social Affairs (2018) World urbanization prospects. technical report United Nations Department of Economic and Social Affairs
- Venigalla M, Kaviti S, Brennan T (2020) Impact of bikesharing pricing policies on usage and revenue: an evaluation through curation of large datasets from revenue transactions and trips. *J Big Data Anal Transp* 2:1–16
- Weigl S, Bogenberger K (2013) Relocation strategies and algorithms for free-floating car sharing systems. *IEEE Intell Transp Syst Mag* 5:100–111. <https://doi.org/10.1109/MITS.2013.2267810>
- Wendland H (2004) Scattered data approximation, 1st edn. Cambridge University Press. <https://doi.org/10.1017/CBO9780511617539>
- Wessel J (2020) Using weather forecasts to forecast whether bikes are used. *Transp Res Part A Policy Pract* 138:537–559. <https://doi.org/10.1016/j.tra.2020.06.006>
- Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Philip SY et al (2008) Top 10 algorithms in data mining. *Knowl Inf Syst* 14:1–37
- Xin L, Tianyun S, Xiaoning M (2020) Research on the big data platform and its key technologies for the railway locomotive system. In: Proceedings of the 2020 5th international conference on big data and computing, pp 6–12
- Xu C, Ji J, Liu P (2018) The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transp Res Part C Emerg Technol* 95:47–60
- Yang Z, Hu J, Shu Y, Cheng P, Chen J, Moscibroda T (2016) Mobility modeling and prediction in bike-sharing systems. In: Proceedings of the 14th annual international conference on mobile systems, applications, and services, pp 165–178
- Yang H, Ma Q, Wang Z, Cai Q, Xie K, Yang D (2020a) Safety of micro-mobility: analysis of e-scooter crashes by mining news reports. *Accid Anal Prev* 143:105608
- Yang Y, Heppenstall A, Turner A, Comber A (2020b) Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems. *Comput Environ Urban Syst* 83:101521
- Yoon T, Cherry CR, Jones LR (2017) One-way and round-trip carsharing: a stated preference experiment in Beijing. *Transp Res Part D Transp Environ* 53:102–114
- Younes H, Zou Z, Wu J, Baiocchi G (2020) Comparing the temporal determinants of dockless scooter-share and station-based bike-share in Washington, DC. *Transp Res Part A Policy Pract* 134:308–320
- Zannat KE, Choudhury CF (2019) Emerging big data sources for public transport planning: a systematic review on current state of art and future research directions. *J Indian Inst Sci* 99:601–619
- Zhang J, Zheng Y, Qi D, Li R, Yi X (2016) DNN-based prediction model for spatio-temporal data. In: Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems. ACM Press, Burlingame, pp 1–4. <https://doi.org/10.1145/2996913.2997016>
- Zhang C, He J, Liu Z, Xing L, Wang Y (2019) Travel demand and distance analysis for free-floating car sharing based on deep learning method. *PLoS ONE* 14:e0223973
- Zhang Z, Wang C, Gao Y, Chen J, Zhang Y (2020) Short-term passenger flow forecast of rail transit station based on mic feature selection and st-lightgbm considering transfer passenger flow. *Sci Program* 2020:1–15
- Zhu L, Yu FR, Wang Y, Ning B, Tang T (2018) Big data analytics in intelligent transportation systems: a survey. *IEEE Trans Intell Transp Syst* 20:383–398
- Zhu R, Zhang X, Kondor D, Santi P, Ratti C (2020) Understanding spatio-temporal heterogeneity of bike-sharing and scooter-sharing mobility. *Comput Environ Urban Syst* 81:101483
- Zou Z, Younes H, Erdoğan S, Wu J (2020) Exploratory analysis of real-time e-scooter trip data in Washington, DC. *Transp Res Rec* 2674:285–299

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.