

Shared Micro-mobility Demand Forecasting using Gradient Boosting methods

Antonios Tziorvas^{1,*}, George S. Theodoropoulos^{1,†} and Yannis Theodoridis^{1,†}

¹Department of Informatics, University of Piraeus, Piraeus, Greece

Abstract

Urban demand forecasting plays a critical role in optimizing routing, dispatching, and congestion management within Intelligent Transportation Systems. By leveraging data fusion and analytics techniques, traffic density estimation serves as a key intermediate measure for identifying and predicting emerging demand patterns. In this paper, we propose two gradient boosting model variations, one for classification and one for regression, both capable of generating demand forecasts at various temporal horizons, from 5 minutes up to one hour. Our approach effectively integrates spatial and temporal features, enabling accurate predictions that are essential for improving the efficiency of shared (micro-)mobility services. To evaluate the effectiveness of our approach, we utilize open shared mobility data derived from e-scooters and e-bikes networks in two Dutch metropolitan areas. These real-world datasets enable us to validate our approach and demonstrate its effectiveness in capturing the complexities of modern urban mobility. Ultimately, our methodology offers novel insights on urban micro-mobility management, helping to tackle the challenges arising from rapid urbanization and thus, contributing to more sustainable, efficient, and liveable cities.

Keywords

Gradient Boosting, Demand Forecasting, Spatial/Temporal Features, Open shared mobility data, E-scooters/E-bikes, Urban micro-mobility management, Intelligent Transportation Systems

1. Introduction

Urban shared mobility, also known as Mobility-as-a-Service (MaaS) [1], integrates transportation modes, such as public transit, micromobility services (e.g., bike- and scooter-sharing), and commute-based models (e.g., carpooling). The need for accurate mobility pattern forecasting is growing rapidly as real-time data analytics in Intelligent Transportation Systems (ITS) help alleviate congestion, reduce travel times, and enhance road safety in increasingly complex urban environments. In shared mobility services, including ride-hailing and bike-sharing, predicting demand across spatial and temporal dimensions is essential for efficient resource allocation, reduced waiting times, and optimized fleet deployment. This insight supports smart city initiatives by informing human-centric urban infrastructure design and facilitating sustainable development through data-driven decisions in areas, such as energy consumption, public transit planning, and emergency services.

Several Machine Learning (ML) approaches have been proposed for detecting and forecasting spatio-temporal patterns in timeseries, including Long Short-Term Memory (LSTM) networks [2], Graph Neural Networks (GNNs) [3, 4], and Diffusion-based Models [5, 6]. However, these models often suffer from high computational complexity due to the intricate nature of spatio-temporal data, which combines spatial correlations with temporal dependencies [7].

A common representation for such data is the Spatio-Temporal Graph (STG), where nodes denote spatial locations and edges encode relationships over time. Capturing both spatial and temporal dependencies requires advanced architectures, such as Spatio-Temporal Graph Neural Networks (STGNNs) [8] or Spatio-Temporal Graph Convolutional Networks (STGCNs) [9]. However, these methods impose sig-

nificant computational demands, limiting their real-world applicability. [10]. While optimization techniques such as compression and pruning can reduce model size and inference time, they often come at the cost of decreased predictive accuracy.

This trade-off between complexity, accuracy, and efficiency underscores the need for alternative methodologies. In this work, we propose a novel approach based on gradient-boosted trees, a class of models known for their superior performance on structured tabular data [11]. Our method effectively forecasts micro-mobility demand across spatial and temporal dimensions while maintaining computational efficiency, making it suitable for large-scale deployment.

In our approach, we provide a robust feature extraction pipeline that can capture both spatial and temporal dependencies and integrate them into our model. We also present a gradient boosting ML algorithm capable of predicting the demand in a given area, either in the form of levels (e.g., 'Low', 'Medium', 'High') through a classifier, or in absolute demand values through a regressor. The performance of our model is evaluated using two real-world micro-mobility datasets and the results turn out to be quite promising since the our approach can effectively adapt to the uniqueness of each area and adequately model its intricacies.

The rest of this paper is structured as follows: Section 2 reviews related work, Section 3 introduces our forecasting methodology, including the feature extraction process and training configuration. Section 4 describes the experimental setup and presents the results of both predictive models, offering detailed performance metrics. Finally, Section 5 concludes the paper with a summary of the key contributions and directions for future research.

2. Related Work

A wide variety of modeling approaches have been proposed to tackle the challenges related to spatio-temporal forecasting. In this section, we will identify and review previous approaches in the context of shared mobility demand forecasting by categorizing them in three groups: (traditional or advanced) ML approaches, deep learning models, and

Published in the Proceedings of the Workshops of the EDBT/ICDT 2025 Joint Conference (March 25-28, 2025), Barcelona, Spain

* Corresponding author.

† These authors contributed equally.

✉ atzio@unipi.gr (A. Tziorvas); gstheo@unipi.gr

(G. S. Theodoropoulos); ytheod@unipi.gr (Y. Theodoridis)

🆔 0009-0005-0037-6264 (A. Tziorvas); 0000-0003-4547-6646

(G. S. Theodoropoulos); 0000-0003-2589-7881 (Y. Theodoridis)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ensemble and hybrid methods, respectively.

2.1. ML Approaches

Traditional machine learning (ML) methods provide a stable analytical foundation for structured predictive problems. Feng et al. [12] developed a predictive model for bike-sharing demand in Chicago using Poisson regression, incorporating time, weather, population density, and activity density as features. Additionally, a Random Forest model was employed to enhance predictive accuracy through ensemble learning, aiding in demand-supply optimization and the identification of potential new station locations. Similarly, Xiao et al. [13] applied clustering techniques to forecast short-term car demand, utilizing an improved k-nearest neighbor (kNN) model for online car-hailing services in Hefei City, China. Their results highlight the effectiveness of clustering methodologies in enhancing demand prediction accuracy.

A notable advancement in bike-share traffic prediction is the two-step pattern recognition model proposed by Sohrabi & Ermagun [14]. This approach first generates traffic profiles, where bike station traffic is represented as timeseries profiles with t -minute intervals and an overlap length O . The kNN algorithm is then applied to identify spatiotemporal similarities, incorporating historical traffic data, temporal factors (e.g., weather, weekdays), and spatial characteristics (e.g., socio-demographics, land use, infrastructure). By leveraging weighted Euclidean distance metrics, this method enables accurate short- and long-term traffic forecasting at both station and system-wide levels, demonstrating the efficacy of integrating spatiotemporal features with ML techniques.

Smith et al. [15] employ ML to predict bicycle and pedestrian traffic patterns across 20 U.S. metropolitan statistical areas (MSAs). Building on previous efforts by Le et al. [16], where stepwise linear regression models were developed based on Census-derived, neighborhood-level covariates, this study incorporates traffic counts aggregated from 4,145 locations alongside novel street-level data, including Google Street View imagery and Point of Interest (PoI) data. The authors evaluate several ML regressors, including linear, ridge, lasso, bagging, gradient boosting, and random forest, reporting significant improvements in predictive accuracy over traditional linear regression, particularly when using augmented datasets.

2.2. Deep Learning Models

Ge et al. [17] introduced a Self-Attention ConvLSTM (SA-ConvLSTM) model to enhance the accuracy of ConvLSTM for online car-hailing demand forecasting. By converting car-hailing trajectories into grid-based images, the model incorporates a self-attention module to capture long-range spatiotemporal dependencies, leveraging pairwise similarity scores across input and memory positions.

Luo et al. [18] proposed a Spatial-Temporal Diffusion Convolutional Network (ST-DCN) to address limitations in modeling dynamic spatiotemporal dependencies for taxi demand forecasting. Their approach integrates a two-phase graph diffusion convolutional network with an attention mechanism to model spatial dependencies, while a temporal convolution module captures long-term trends, including recent, daily, and weekly patterns. The use of stacked convolution layers further enhances the model's ability to process

extended timeseries sequences.

Li et al. [19] proposed a hierarchical framework for proactive bike redistribution in bike-sharing systems. A bipartite clustering algorithm groups stations, followed by city-wide rental prediction using a Gradient Boosting Regression Tree. Rental proportions and inter-cluster transitions are then estimated via a multi-similarity-based model to predict station-level rentals and returns. Experiments on real-world datasets from New York City and Washington D.C. indicate significant performance gains over baseline models, particularly during periods of atypical demand.

2.3. Ensemble and Hybrid Models

The concept of ensemble learning, which involves combining the predictions of multiple simple models to form a new, more robust model, has been extensively explored in the literature and has been shown to produce better results compared to its individual constituent models. Building on this idea, Yuming et al. [20] developed a stacking ensemble learning framework that integrates the predictions of three distinct base learners: Random Forest, LightGBM, and Long Short-Term Memory (LSTM) networks. These base learners were trained in parallel using a common set of input features, including temporal, spatial, and weather-related variables. The predictions generated by each base learner were then aggregated by a Support Vector Regression (SVR) meta-learner that produced the final demand forecast values.

2.4. Positioning of Our Work with Respect to State-of-the-Art

Our approach diverges from these approaches in multiple ways. Firstly, our methodology employs a unified modeling workflow that can tackle the problem in two alternative ways, by either regression or classification, allowing us to adjust to specific needs. Additionally, our feature engineering pipeline integrates both spatial and temporal features which are fed to a tuned gradient boosting model, enabling us to better identify the intricacies of urban micro-mobility. Lastly, by utilizing two real-world datasets for evaluation, our method can be adequately validated under different scenarios, ensuring it is effective and robust. Nevertheless, as part of our future research, we plan to align related work with our approach for a fair comparison under the same settings over the same real datasets.

3. Our Approach for Efficient Shared Micro-Mobility Demand Forecasting

In this section, we define the shared mobility demand forecasting problem and present our proposed approach.

3.1. Problem Definition

Generally speaking, a spatio-temporal timeseries is a sequence of data points, where each one is associated with a time index and a spatial location. For the purposes of our work, an area of interest (a city, a region, etc.) is partitioned into a set of geographical districts, where each district A is represented as a polygon. We denote D_t^A as

the percentile-encoded demand of micro-mobility vehicles within district A at time t . Given τ past observations for district A , we construct its spatio-temporal timeseries as $X_t^A = \{D_{t-\tau}^A, D_{t-\tau+1}^A, \dots, D_{t-1}^A, D_t^A\}$. The problem addressed in this paper is to predict the micro-mobility demand value D_{t+m}^A for district A , m time steps ahead, given X_t^A i.e., the current along with the past τ observations.

We will exemplify the above definition through a real case extracted from our real-world dataset (dataset details are presented in Section 4). As illustrated in Figure 1, the problem at hand is to predict (as accurately as possible) the anticipated demand values of each district’s demand timeseries (in orange) given the current and some past demand values (in blue), taking into account relevant (temporal, spatial, context) features. To efficiently achieve this goal, seasonal etc. patterns hidden in these timeseries should be disclosed. For instance, looking at the timeseries of two major districts in Rotterdam that are displayed in Figure 1, we identify seasonal commuting patterns: the highest and most spread peak occurs on Friday, due to a combination of leisure and work time. This is made more evident in Figure 2, which showcases the combination of the hour-of-day and day-of-week features in Rotterdam Centrum district.

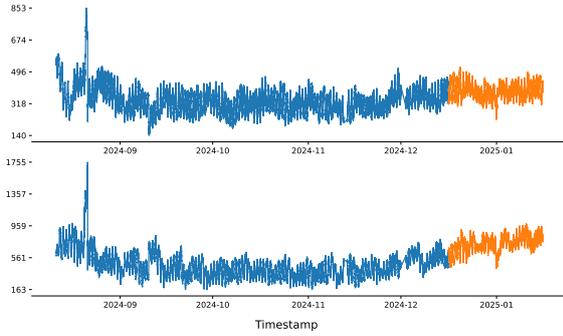


Figure 1: Two indicative timeseries for two regions of Rotterdam: Delfshaven (top) and Rotterdam Centrum (bottom). The x-axis denotes time, while the y-axis indicates the measured demand values.

3.2. Methodology

We propose the so-called *Gradient Boosting Demand Predictor*, in two variations, differentiated by their output type:

- A N -class classifier (per horizon per district)
- A regressor (per horizon per district)

The purpose of the classifier is to classify the future demand of a district into predefined levels (e.g., ‘Low’, ‘Medium’, ‘High’), whereas the regressor aims to predict the actual demand values for the respective horizon. In other words, for an area of interest partitioned into M districts with the goal of T individual prediction horizons, both architectures require training $T \times M$ models, where each model predicts the output for a specific district-horizon pair. The reasoning behind training individual models per district is to allow each model to specialize in its respective spatial region, focusing exclusively on its unique spatial and temporal patterns. Nevertheless, in our future work we plan to assess the efficiency of single global model for all districts (where the district ID will be given as input to the model).

As illustrated in Figure 3, the initial step of our methodology is the partitioning of the area of interest into districts,

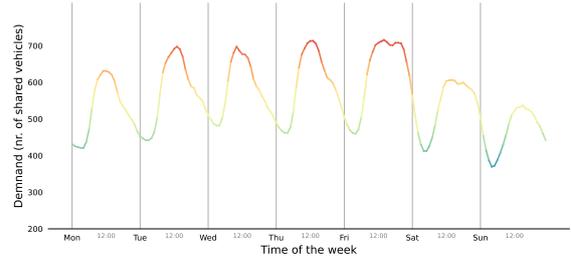


Figure 2: Average hourly demand across the week in Rotterdam Centrum. The coloring indicates the relative demand intensity and, along with the peaks, aid in highlighting daily and weekly usage patterns.

modeled as an adjacency graph that allows us to extract the neighbors of each district. Two districts are defined as *neighbors* if they share a common boundary¹. The demands of the target as well as of the neighboring districts are represented by the respective spatiotemporal timeseries (step 2), as defined in Section 3.1, which are fed into a feature extraction pipeline (step 3), to be detailed in Section 3.3. Finally, all these features are combined into a single, augmented dataset (step 4) to be used for the purposes of training our two models, the classifier and the regressor (step 5).

3.3. Feature Extraction

Given a timeseries X_t^A of a target district A , feature extraction is performed to enhance the spatio-temporal modeling capabilities of the proposed method. The extracted features are categorized into four primary groups: time-related, lagged, rolling, and exponentially weighted features. These features are derived not only from the target district but also from its neighboring districts, incorporating both spatial and temporal dependencies.

Time-related capture temporal patterns influencing the target variable. Key attributes include the hour of the day (0-23) to identify daily patterns like peak periods, and the day of the week (0-6, with 0 as Monday) to capture weekly trends. Monthly variations are represented by the month (0-11, with 0 as January), accounting for seasonal behaviors. Additionally, the minute component (0-59) highlights finer time dependencies and intra-hour variations.

The second category of features, denoted to as *lagged* features, is designed to capture the temporal dependencies between past observations of the timeseries and its current state. For each district, a set of lagged features is generated by systematically shifting the original timeseries observations backward in time. Formally, given a timeseries X_t^A representing the observed values of a target district A at time t , the lagged feature at a temporal offset τ is defined as $X_{t-\tau}^A$. A range of temporal offsets is selected to ensure that both short-term and long-term dependencies are captured. Specifically, the lagged features are computed at intervals of 1, 5, 10, and 15 minutes, capturing short-term temporal dependencies across multiple time horizons.

¹The inclusion of neighboring districts is justified by the limited spatial range and temporal duration of trips usually made by e-bikes and e-scooters. Due to their physical constraints, micro-mobility vehicles are unlikely to reach distant districts within short timeframes, making the consideration of neighboring districts only being a sufficient choice for effective spatio-temporal modeling.

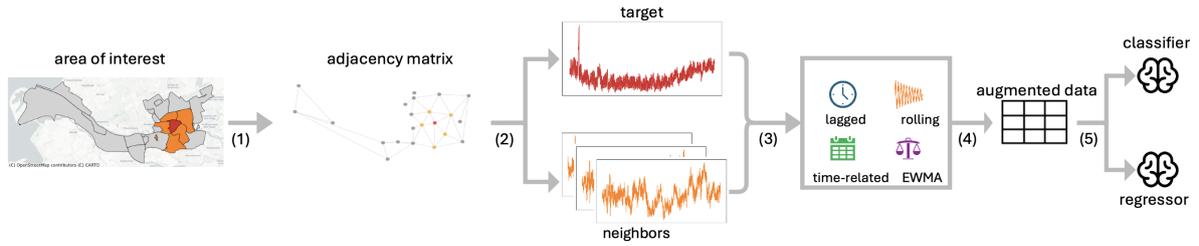


Figure 3: Illustration of our Gradient Boosting Density Predictor proposed methodology

Table 1
Summary table of the extracted features used in our approach

Category	Feature Name	Description
Base	target	Target variable for the current district at the present time step
	neighbor_*	Feature variable of each of the target’s neighbors in neighboring districts
Time-Related	hour-of-day	Hour of the day of the timestamp (0-23)
	minute-of-hour	Minute of the hour of the timestamp (0-59)
	day-of-week	Day of the week of the timestamp (0=Monday, 6=Sunday)
	month-of-year	Month of the year of the timestamp (0=January, 11=December)
Lagged	neighbor_lag_{1, 5, 10, 15}	Lagged value ({1, 5, 10, 15}-step(s) back) of the target variable for neighboring districts
	target_lag_{1, 5, 10, 15}	Lagged value ({1, 5, 10, 15}-step(s) back) of the target variable for the current district
Rolling	neighbor_rolling_mean_{5, 10, 15}	Rolling mean of neighbors over the last {5, 10, 15} observations
	target_rolling_mean_{5, 10, 15}	Rolling mean of the target variable over the last {5, 10, 15} observations
EWMA	neighbor_ewm_{5, 10, 15}	Exponentially weighted moving average of neighbors with a span of {5, 10, 15}
	target_ewm_{5, 10, 15}	Exponentially weighted moving average of the target variable with a span of {5, 10, 15}

This shifting operation effectively creates additional feature representations of the past states of the system, allowing the model to learn historical patterns that influence future predictions. By incorporating lagged features into the predictive framework, the approach enhances its ability to capture recurrent patterns, trends, and autocorrelations inherent in spatio-temporal mobility data.

The *rolling* features provide a smoothed representation of the data over various intervals, capturing localized fluctuations that are critical for accurate forecasting. These features are computed using sliding windows, where statistical measures are aggregated over a sequence of recent time steps to encapsulate short-term trends and variability in mobility patterns. For each district, a rolling window of varying sizes—ranging from 5 to 15 minutes in 5-minute increments—is applied to the timeseries data. Within each window, the mean is computed to capture central tendency and, over larger windows, the degree of short-term variability.

Finally, the objective of the *exponentially weighted* features is to prioritize recent observations while retaining information from historical data. This is achieved through the application of exponentially weighted techniques, which assign progressively smaller weights to older data points, ensuring that the model places greater emphasis on more recent trends. Among these methods, the Exponential Weighted Moving Average (EWMA) is particularly effective, as it assigns exponentially decreasing weights to past observations, thereby capturing temporal dependencies in a manner that balances adaptability and historical context. For a given timeseries X_t , the EWMA at time t is computed recursively as $EWMA_t = \alpha X_t + (1 - \alpha)EWMA_{t-1}$ where α represents the smoothing factor, typically defined as $\alpha = \frac{2}{N+1}$ for a given window size N . This formulation allows the predictive framework to swiftly adapt to

evolving patterns, including sudden variations induced by external factors such as weather conditions, road incidents, or public events, while maintaining a stable representation of long-term trends.

The features used in our methodology are summarized in Table 1.

4. Experimental Study

This section describes the data and preprocessing steps used in our experimental study (Section 4.1) as well as our findings (Section 4.2). The experiments were conducted on a server with an AMD Epyc 64-Core CPU and an Nvidia A100 GPU with 40GB of memory. Our source code is publicly available for reproducibility purposes.²

4.1. Experimental Setup

In our study, we utilize shared micro-mobility data provided by Deelfiets Nederland, a popular micro-mobility service provider in Netherlands, through an open API³. Data consists of real-time GPS positions for vehicles in the Netherlands, which we query every 60 seconds, creating a dataset of sufficient granularity for our use-case. The dataset includes the latitude, longitude, vehicle type and company name for each vehicle. For the purposes of this study, we utilize mobility data from two major metropolitan areas, Amsterdam and Rotterdam, focusing specifically on the ten most densely populated districts in each city. These districts exhibit diverse urban characteristics and distinct spatio-temporal patterns, making them well-suited for evaluating the model’s adaptability to localized mobility trends. The

²<https://github.com/DataStories-UniPi/Shared-Mobility.git>

³https://api.deelfietsdashboard.nl/dashboard-api/public/vehicles_in_public_space

temporal coverage of our data is about 2 months (November 11, 2024 - January 15, 2025), and 5 months (August 06, 2024 - January 15, 2025), respectively. Since the original data source is a stream instead of a static dataset, we manually append the temporal indicator in each subsequent request.

To derive demand estimates at the district level, a spatial filtering and aggregation procedure is applied. First, a point-polygon intersection is performed to associate each mobility trace with its corresponding district boundaries, discarding traces that fall outside the predefined study areas. Next, the filtered traces are aggregated based on their unique spatio-temporal identifiers (i.e., timestamp and district ID), summing the occurrences within each region to obtain demand estimates. This transformation preserves the spatial and temporal integrity of the data while ensuring compatibility with the forecasting models. In its final form, the dataset consists of 89,022 observations across 97 districts in Amsterdam and 218,325 observations across 22 districts in Rotterdam.

To validate our models, we use a 70-30 train-test temporal split validation method. To ensure effective training of the classifier, additional preprocessing steps were taken apart from the feature extraction process discovered above. To this extent, a quantile-based discretization function was employed to transform continuous density values into 3 discrete demand levels (Low, Medium, High, as they will be defined in the following paragraph). This transformation enhances the classifier’s ability to model class distinctions and mitigates the challenges associated with imbalanced class distributions.

We define ‘Medium’ demand as the range where demand levels are centered around 50% of the value of the day with the highest demand, which serves as the baseline value. Specifically, for a given threshold $d \in (0, 50)$, the range $[50 - d, 50 + d]$ is categorized as normal demand. Demand levels below (above) this range are classified as ‘Low’ (‘High’, respectively) demand. In our experiments, we set $d = 17$, in order for the demand levels to correspond to three equally sized ranges. This threshold provides a balanced distribution of data across the defined demand categories, facilitating a robust analysis of varying demand patterns.

The optimization framework employed in this study incorporates Bayesian Optimization due to its ability to balance exploration and exploitation effectively. This approach facilitates the exploration of the search space for better solutions while simultaneously exploiting regions with high potential, thereby reducing the number of iterations required for optimization. Its suitability is particularly evident when dealing with extensive experimental evaluations, such as the computationally intensive training of ML models, as it leverages prior knowledge to guide sampling and optimize resources efficiently. Additionally, the probabilistic framework of Bayesian Optimization enables the automatic adjustment of the exploration-exploitation trade-off through its acquisition functions.

4.2. Experimental Results

In this section, we assess the overall performance of our models across multiple districts and timestamps. Table 2 summarizes the training and inference times of our models across different prediction horizons in the two cities of interest. Training time corresponds to the overall time required to train the respective model, whereas inference time corresponds to the average time taken to generate a

specific prediction for a given time horizon. It turns out that training of our models is extremely fast since it takes 1-2 seconds or even less to be performed, while the inference time is impressively in the order of 1 microsecond. This analysis showcases the computational efficiency of the models, providing insights into their scalability and suitability for real-world spatio-temporal forecasting tasks.

Table 2

Training and inference times of the two models across different horizons.

City	Model	Prediction Horizon (min)	Training Time (sec)	Inference Time (usec)
Rotterdam	Regressor	5	1.17 ± 0.43	0.99 ± 0.15
		15	0.35 ± 1.04	1.07 ± 0.30
		30	0.22 ± 0.97	0.97 ± 0.09
		60	0.47 ± 1.10	1.10 ± 0.24
	Classifier	5	2.21 ± 1.26	1.11 ± 0.17
		15	1.10 ± 1.15	1.15 ± 0.21
Amsterdam	Regressor	30	1.00 ± 1.10	1.10 ± 0.18
		60	1.07 ± 1.19	1.19 ± 0.16
		5	0.68 ± 0.22	0.95 ± 0.12
		15	1.21 ± 0.95	0.95 ± 0.18
	Classifier	30	0.28 ± 0.94	0.94 ± 0.13
		60	0.29 ± 0.92	0.92 ± 0.08
Amsterdam	Regressor	5	1.40 ± 0.69	1.01 ± 0.13
		15	0.81 ± 1.10	1.10 ± 0.26
		30	1.22 ± 1.07	1.07 ± 0.28
		60	1.07 ± 1.02	1.02 ± 0.21
	Classifier	5	1.40 ± 0.69	1.01 ± 0.13
		15	0.81 ± 1.10	1.10 ± 0.26

Tables 3-6 present the predictive performance of the proposed spatio-temporal forecasting models evaluated at four prediction horizons: 5, 15, 30, and 60 minutes. The quality metrics we used include R^2 , RMSE, sMAPE and MASE for the regressor and F1-score, Accuracy, Recall and Precision for the classifier. These metrics collectively assess the model’s predictive accuracy, robustness, and relative performance across districts with varying population sizes and temporal dynamics.

Aggregated metrics are added to summarize the overall performance across all districts, offering insights into the model’s general effectiveness and consistency across varying prediction horizons. It is evident that both architectures perform really well at both short-term (5-15 min.) and long-term (30-60 min.) forecasting. From the summary statistics we can deduce that both models consistently perform well for the vast majority of the districts. Indicatively, the regressor’s R^2 ranges on the average from 0.94 down to 0.84 as we move from lower to higher prediction horizons. As for the classifier, $F1$ -score ranges on the average from 0.89 to 0.80, respectively.

4.3. A Note on Feature Importance

In order to assess the effect of the underlying features in the quality of prediction of either the classifier or the regressor, we present the average feature importance across all prediction horizons for both models. The feature importance values were computed using XGBoost’s internal feature importance computation algorithm. Given the minimal contributions of some features (importance values below 10^{-4}), the y-axis of the corresponding plots has been log-scaled to improve visualization. The results of this analysis are grouped by city and illustrated in Figures 4 and 5.

In both figures we can see that the average contribution of the features related to the target is higher than those of the neighbors, as expected. Overall, in decreasing order, the most important features appear to be the *target_ewm_**, the *target_rolling_mean_**, and the *target_lag_**, denoting the EWMA, the rolling mean, and the lagged value, respectively,

Table 3
Predictive performance of the regressor in Rotterdam across different prediction horizons.

District	Avg Pop	5 min				15 min				30 min				60 min			
		R ²	RMSE	MAE	sMAPE	R ²	RMSE	MAE	sMAPE	R ²	RMSE	MAE	sMAPE	R ²	RMSE	MAE	sMAPE
Kralingen-Crooswijk	507	0.97	9.70	6.94	0.01	0.96	11.00	7.86	0.02	0.95	13.17	9.81	0.02	0.90	18.11	13.45	0.03
Rotterdam Centrum	502	0.99	15.25	11.18	0.02	0.98	20.33	14.77	0.02	0.97	22.95	17.37	0.03	0.93	34.66	26.62	0.04
Hillegersberg-Schiebroek	476	0.80	17.36	12.67	0.02	0.76	18.82	14.06	0.03	0.73	20.03	15.20	0.03	0.68	21.79	16.78	0.03
Delfshaven	347	0.98	6.43	4.70	0.01	0.97	8.03	5.75	0.02	0.96	10.13	7.27	0.02	0.92	13.78	9.97	0.03
Feijenoord	347	0.97	6.31	4.61	0.01	0.92	10.09	7.72	0.02	0.94	9.29	7.00	0.02	0.92	10.31	7.63	0.02
Noord	307	0.88	15.69	10.11	0.03	0.85	17.49	11.61	0.03	0.81	19.76	13.61	0.04	0.73	23.96	16.87	0.04
Prins Alexander	303	0.99	6.04	4.37	0.01	0.99	5.91	4.26	0.01	0.98	6.76	4.92	0.01	0.98	7.99	5.87	0.02
Charlois	242	0.97	5.24	3.84	0.02	0.97	5.93	4.42	0.02	0.96	6.89	5.20	0.02	0.94	7.84	5.92	0.03
IJsselmonde	184	0.94	5.25	3.25	0.02	0.90	6.58	4.18	0.03	0.89	7.11	4.50	0.03	0.84	8.37	5.57	0.04
Spaanse Polder	36	0.90	2.18	1.36	0.03	0.69	3.88	2.18	0.05	0.87	2.53	1.76	0.04	0.54	4.75	3.43	0.07
Average ($\mu \pm \sigma$)		0.94 ± 0.06	8.94 ± 5.29	6.30 ± 3.77	0.02 ± 0.01	0.90 ± 0.10	10.80 ± 5.98	7.68 ± 4.41	0.02 ± 0.01	0.90 ± 0.08	11.86 ± 6.86	8.66 ± 5.17	0.03 ± 0.01	0.84 ± 0.14	15.16 ± 9.40	11.21 ± 7.22	0.03 ± 0.02

Table 4
Predictive performance of the classifier in Rotterdam across different prediction horizons.

District	Avg Pop	5 min				15 min				30 min				60 min			
		F1	Acc.	Prec.	Rec.												
Kralingen-Crooswijk	507	0.92	0.92	0.92	0.92	0.90	0.90	0.90	0.90	0.89	0.88	0.88	0.88	0.85	0.85	0.85	0.85
Rotterdam Centrum	502	0.91	0.99	0.96	0.87	0.87	0.98	0.94	0.80	0.83	0.98	0.92	0.78	0.75	0.97	0.84	0.70
Hillegersberg-Schiebroek	476	0.75	0.95	0.82	0.71	0.73	0.94	0.79	0.70	0.71	0.94	0.77	0.67	0.67	0.93	0.76	0.63
Delfshaven	347	0.91	0.96	0.91	0.91	0.89	0.95	0.90	0.89	0.86	0.93	0.86	0.87	0.80	0.90	0.81	0.80
Feijenoord	347	0.88	0.95	0.90	0.87	0.87	0.94	0.89	0.85	0.84	0.93	0.88	0.81	0.74	0.90	0.82	0.71
Noord	307	0.82	0.99	0.85	0.80	0.71	0.98	0.74	0.69	0.55	0.91	0.55	0.63	0.59	0.97	0.62	0.58
Prins Alexander	303	0.97	0.98	0.97	0.97	0.97	0.98	0.97	0.97	0.96	0.98	0.96	0.96	0.96	0.97	0.96	0.95
Charlois	242	0.91	0.91	0.92	0.91	0.90	0.90	0.90	0.90	0.89	0.89	0.89	0.89	0.87	0.87	0.87	0.87
IJsselmonde	184	0.92	0.97	0.92	0.92	0.91	0.97	0.91	0.91	0.90	0.97	0.90	0.91	0.89	0.97	0.89	0.90
Spaanse Polder	36	0.95	0.99	0.98	0.92	0.90	0.98	0.94	0.87	0.93	0.98	0.96	0.90	0.91	0.98	0.95	0.88
Average ($\mu \pm \sigma$)		0.89 ± 0.06	0.96 ± 0.03	0.91 ± 0.05	0.88 ± 0.07	0.87 ± 0.08	0.95 ± 0.03	0.89 ± 0.07	0.85 ± 0.09	0.84 ± 0.12	0.94 ± 0.04	0.86 ± 0.12	0.83 ± 0.11	0.80 ± 0.11	0.93 ± 0.05	0.84 ± 0.10	0.79 ± 0.12

of the target variable. Focusing on the time-related features only, *hour-of-day* and *day-of-week* show the relatively highest importance in both cases, underlying the presence of seasonal patterns.

5. Conclusions

This study investigated spatio-temporal forecasting methods for density prediction in urban environments. By integrating spatial dependencies with temporal trends, the pro-

Table 5
Predictive performance of the regressor in Amsterdam across different prediction horizons.

District	Avg. Pop	5 min				15 min				30 min				60 min			
		R ²	RMSE	MAE	sMAPE	R ²	RMSE	MAE	sMAPE	R ²	RMSE	MAE	sMAPE	R ²	RMSE	MAE	sMAPE
Oostelijk Havengebied	110	0.97	2.34	1.83	0.02	0.96	2.75	2.16	0.02	0.95	3.24	2.53	0.02	0.92	4.16	3.24	0.03
Buitenveldert-West	87	0.97	2.12	1.57	0.02	0.95	2.66	1.96	0.02	0.93	3.28	2.40	0.02	0.87	4.37	3.17	0.03
Burgwallen-Nieuwe Zijde	72	0.98	1.91	1.41	0.02	0.96	2.55	1.91	0.03	0.93	3.38	2.58	0.03	0.89	4.35	3.35	0.04
Jordaan	71	0.98	2.31	1.78	0.02	0.94	4.04	2.75	0.03	0.91	4.67	3.42	0.04	0.85	6.14	4.66	0.05
Scheldebuurt	71	0.98	1.56	1.19	0.01	0.96	2.06	1.57	0.02	0.94	2.66	2.02	0.02	0.89	3.50	2.64	0.03
Middenmeer	66	0.96	1.94	1.46	0.02	0.94	2.28	1.73	0.02	0.90	2.89	2.20	0.03	0.86	3.55	2.71	0.03
IJburg West	63	0.93	1.62	1.28	0.02	0.91	1.83	1.45	0.02	0.87	2.15	1.70	0.02	0.79	2.76	2.20	0.03
Westelijk Havengebied	63	0.88	3.72	1.60	0.03	0.85	4.05	1.75	0.03	0.83	4.28	1.87	0.04	0.78	4.74	2.20	0.04
Landlust	59	0.99	1.42	1.07	0.02	0.97	1.96	1.51	0.03	0.96	2.60	2.01	0.04	0.92	3.49	2.70	0.05
Nieuwmarkt/Lastage	58	0.92	2.82	1.94	0.02	0.88	3.34	2.37	0.03	0.76	4.71	3.35	0.04	0.70	5.29	3.88	0.05
Average ($\mu \pm \sigma$)		0.95 ± 0.04	2.18 ± 0.69	1.51 ± 0.28	0.02 ± 0.01	0.93 ± 0.04	2.75 ± 0.81	1.92 ± 0.41	0.02 ± 0.01	0.90 ± 0.06	3.38 ± 0.89	2.41 ± 0.59	0.03 ± 0.01	0.85 ± 0.07	4.24 ± 0.99	3.08 ± 0.77	0.04 ± 0.01

Table 6
Predictive performance of the classifier in Amsterdam across different prediction horizons.

District	Avg. Pop	5 min				15 min				30 min				60 min			
		F1	Acc.	Prec.	Rec.												
Oostelijk Havengebied	110	0.90	0.92	0.93	0.89	0.89	0.90	0.89	0.88	0.86	0.88	0.87	0.85	0.76	0.78	0.75	0.79
Buitenveldert-West	87	0.92	0.93	0.94	0.91	0.90	0.91	0.92	0.89	0.88	0.89	0.89	0.86	0.83	0.86	0.85	0.82
Burgwallen-Nieuwe Zijde	72	0.94	0.94	0.95	0.94	0.92	0.92	0.92	0.91	0.89	0.89	0.89	0.89	0.85	0.84	0.85	0.84
Jordaan	71	0.91	0.94	0.94	0.89	0.87	0.92	0.89	0.85	0.84	0.90	0.87	0.81	0.82	0.89	0.82	0.82
Scheldebuurt	71	0.91	0.93	0.93	0.90	0.88	0.90	0.90	0.87	0.85	0.88	0.88	0.83	0.79	0.83	0.83	0.77
Middenmeer	66	0.85	0.94	0.90	0.81	0.79	0.91	0.81	0.77	0.73	0.90	0.80	0.70	0.68	0.83	0.69	0.69
IJburg West	63	0.73	0.96	0.81	0.68	0.63	0.95	0.66	0.61	0.57	0.79	0.54	0.73	0.61	0.92	0.60	0.62
Westelijk Havengebied	63	0.91	0.89	0.94	0.90	0.89	0.87	0.93	0.88	0.90	0.89	0.92	0.90	0.87	0.85	0.89	0.86
Landlust	59	0.94	0.94	0.93	0.94	0.93	0.93	0.92	0.93	0.92	0.92	0.91	0.92	0.87	0.87	0.87	0.87
Nieuwmarkt/Lastage	58	0.79	0.94	0.88	0.73	0.69	0.92	0.78	0.65	0.62	0.91	0.65	0.60	0.55	0.88	0.57	0.53
Average ($\mu \pm \sigma$)		0.88 ± 0.07	0.93 ± 0.02	0.91 ± 0.04	0.86 ± 0.09	0.84 ± 0.10	0.91 ± 0.02	0.86 ± 0.09	0.83 ± 0.11	0.81 ± 0.12	0.88 ± 0.03	0.82 ± 0.13	0.81 ± 0.10	0.76 ± 0.11	0.86 ± 0.04	0.77 ± 0.11	0.76 ± 0.11

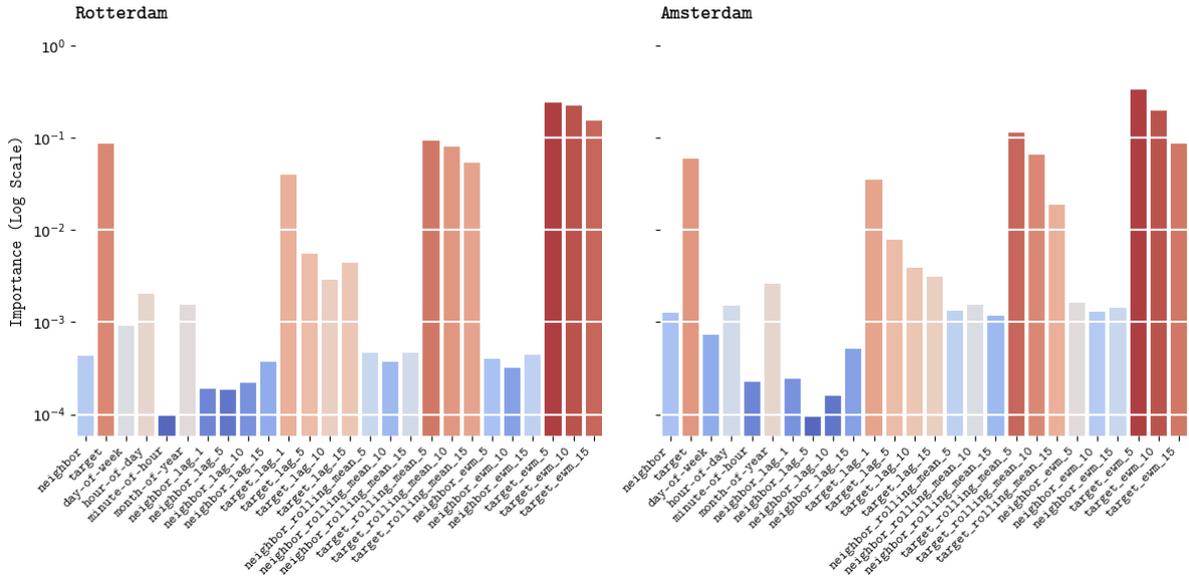


Figure 4: Average feature importance across all 4 prediction horizons of the regressor in Rotterdam (left) and Amsterdam (right)

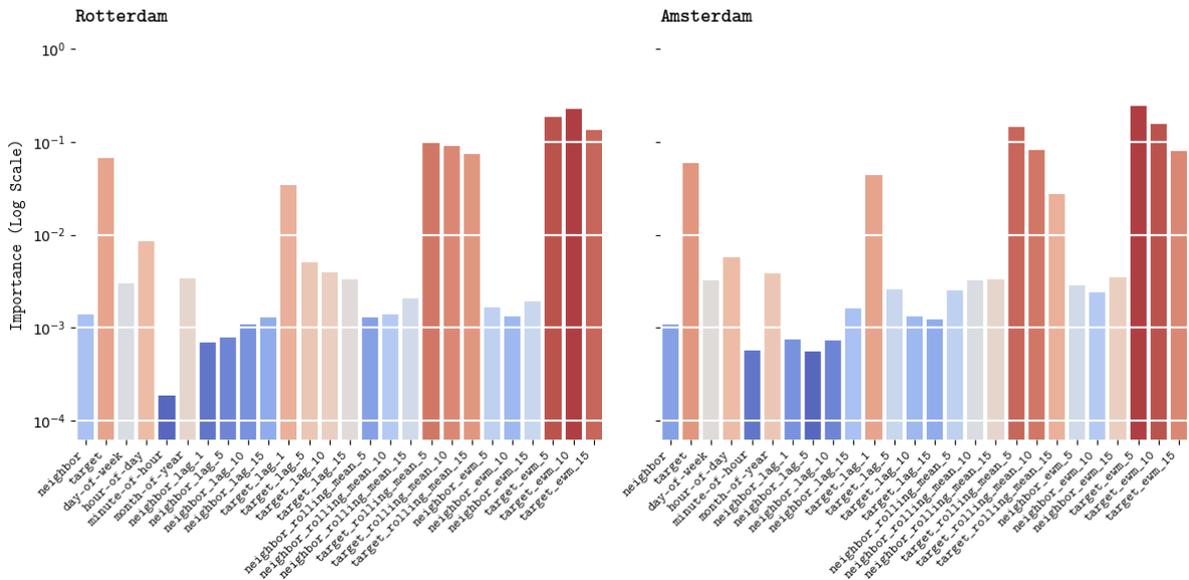


Figure 5: Average feature importance across all 4 prediction horizons of the classifier in Rotterdam (left) and Amsterdam (right)

posed methodology effectively captured localized patterns, generating accurate forecasts across multiple horizons. Key contributions include a flexible feature engineering pipeline that incorporates both intra- and inter-district interactions, alongside the application of efficient gradient boosting architectures. This approach enhances predictive accuracy while minimizing computational overhead, rendering it suitable for large-scale forecasting tasks. Experimental results demonstrated the efficacy of gradient-boosted density predictors for both regression and classification, exhibiting competitive performance across varying prediction horizons. Although accuracy slightly declines for longer horizons, the model remains robust, underscoring its adaptability and practical applicability in real-world mobility forecasting.

This research also unveils opportunities for future exploration. Specifically, the integration of feature interactions and their contributions to the overall model quality warrant further investigation. Incorporating external factors, such as weather conditions or public events, could potentially enhance predictive performance. Additionally, applying the proposed methodology to other geographical areas and diverse mobility scenarios would help validate its generalization capabilities and adaptability to varying contexts. Finally, a comprehensive experimental comparison with related work under specific settings is planned to facilitate fair benchmarking.

In conclusion, this study showcases the ability of data-driven approaches to effectively tackle spatio-temporal fore-

casting challenges. By leveraging the inherent spatial segmentation of cities into districts, the methodology enables the extraction of localized temporal patterns, facilitating more informed decision-making and contributing to smarter, more efficient urban planning from the perspective of mobility data science [21].

Acknowledgments

This work was supported in part by the Horizon Framework Programme of the European Union under grant agreement No. 101093051 (EMERALDS; <https://www.emeralds-horizon.eu/>).

Declaration on Generative AI

During the preparation of this work, the authors used Large Language Models in order to paraphrase and reword. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] M. N. Mladenović, Mobility as a Service, in: R. Vickerman (Ed.), *International Encyclopedia of Transportation*, Elsevier, 2021, pp. 12–18. doi:10.1016/B978-0-08-102671-7.10607-4.
- [2] F. O'Donncha, Y. Hu, et al., A spatio-temporal LSTM model to forecast across multiple temporal and spatial scales, *Ecological Informatics* 69 (2022) 101687. doi:10.1016/j.ecoinf.2022.101687.
- [3] B. Yu, H. Yin, et al., Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting, in: *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018, p. 3634–3640. doi:10.24963/ijcai.2018/505.
- [4] C. Zheng, X. Fan, et al., GMAN: A Graph Multi-Attention Network for Traffic Prediction, *Proceedings of the AAAI Conference on Artificial Intelligence* 34 (2020) 1234–1241. doi:10.1609/aaai.v34i01.5477.
- [5] Y. Yang, M. Jin, et al., A Survey on Diffusion Models for Time Series and Spatio-Temporal Data, *ArXiv abs/2404.18886* (2024). doi:10.48550/arXiv.2404.18886.
- [6] Y. Li, R. Yu, et al., Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting, *arXiv: 1707.01926* (2017). doi:10.48550/arXiv.1707.01926.
- [7] N. Pelekis, Y. Theodoridis, *Mobility Data Management and Exploration*, Springer, New York, NY, 2014. doi:10.1007/978-1-4939-0392-4.
- [8] W. Liao, B. Zeng, et al., Taxi demand forecasting based on the temporal multimodal information fusion graph neural network, *Applied Intelligence* 52 (2022) 12077–12090. doi:10.1007/s10489-021-03128-1.
- [9] X. Geng, Y. Li, et al., Spatiotemporal Multi-Graph Convolution Network for Ride-Hailing Demand Forecasting, *Proceedings of the AAAI Conference on Artificial Intelligence* 33 (2019) 3656–3663. doi:10.1609/aaai.v33i01.33013656.
- [10] Z. Wu, S. Pan, et al., A Comprehensive Survey on Graph Neural Networks, *IEEE Transactions on Neural Networks and Learning Systems* 32 (2019) 4–24. doi:10.1109/TNNLS.2020.2978386.
- [11] L. Grinsztajn, E. Oyallon, et al., Why do tree-based models still outperform deep learning on typical tabular data?, in: *Proceedings of the 36th International Conference on Neural Information Processing Systems, (NIPS '22)*, 2022, pp. 507–520. doi:<https://doi.org/10.48550/arXiv.2207.08815>.
- [12] H. Feng, Predicting the Dynamic Demand of Bike-Sharing System in Chicago with Divvy Operation Data: A Data-Driven approach for bike-sharing demand forecasting, in: *Proceedings of the 5th International Conference on E-Commerce, E-Business and E-Government, ICEEG '21*, 2021, p. 30–34. doi:10.1145/3466029.3466035.
- [13] Y. Xiao, W. Kong, et al., Short-Term Demand Forecasting of Urban Online Car-Hailing Based on the K-Nearest Neighbor Model, *Sensors* 22 (2022). doi:10.3390/s22239456.
- [14] S. Sohrabi, A. Ermagun, Dynamic bike sharing traffic prediction using spatiotemporal pattern detection, *Transportation Research Part D: Transport and Environment* 90 (2021) 102647. doi:10.1016/j.trd.2020.102647.
- [15] S. Hankey, W. Zhang, et al., Predicting bicycling and walking traffic using street view imagery and destination data, *Transportation Research Part D-transport and Environment* 90 (2021) 102651. doi:10.1016/j.trd.2020.102651.
- [16] H. T. K. Le, R. Buehler, et al., Correlates of the Built Environment and Active Travel: Evidence from 20 US Metropolitan Areas, *Environmental Health Perspectives* 126 (2018). doi:10.1289/EHP3389.
- [17] H. Ge, S. Li, et al., Self-Attention ConvLSTM for Spatiotemporal Forecasting of Short-Term Online Car-Hailing Demand, *Sustainability* 14 (2022). doi:10.3390/su14127371.
- [18] A. Luo, B. Shangguan, et al., Spatial-Temporal Diffusion Convolutional Network: A Novel Framework for Taxi Demand Forecasting, *ISPRS Int. J. Geo Inf.* 11 (2022) 193. doi:10.3390/ijgi11030193.
- [19] Y. Li, Y. Zheng, et al., Traffic prediction in a bike-sharing system, in: *Proceedings of the 23rd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Sigspatial '15*, 2015. doi:10.1145/2820783.2820837.
- [20] Y. Jin, X. Ye, et al., Demand Forecasting of Online Car-Hailing With Stacking Ensemble Learning Approach and Large-Scale Datasets, *IEEE Access* 8 (2020) 199513–199522. doi:10.1109/ACCESS.2020.3034355.
- [21] M. Mokbel, M. Sakr, et al., Mobility Data Science: Perspectives and Challenges, *ACM Trans. Spatial Algorithms Syst.* 10 (2024). doi:10.1145/3652158.