Predicting costs of local public bus transport services through machine learning methods

Andrea Amicosante
Affiliation: Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila, L'Aquila, Italy
Email: andrea.amicosante@student.univaq.it

Alessandro Avenali
Affiliation: Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy
Email: alessandro.avenali@uniroma1.it

Tiziana D'Alfonso
Affiliation: Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy
Email: tiziana.dalfonso@uniroma1.it

Mirko Giagnorio
Affiliation: Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy
Email: mirko.giagnorio@uniroma1.it

Andrea Manno
Affiliation: Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila, L'Aquila, Italy
Email: andrea.manno@univaq.it

Giorgio Matteucci
Affiliation: Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy
Email: giorgio.matteucci@uniroma1.it

Abstract

The present study developed several machine learning-based cost models to predict an efficient total economic cost per vehicle revenue-mile of urban public bus transport. The models were trained on a built-in dataset from 269 transit agencies providing urban services in the United States from 2015–2019. A feature selection strategy was implemented, finding that, for each proposed model, a subset of features determined a large impact on unit cost. These “core” features included commercial speed, average salary expenses per employee, vehicle productivity, and fleet ownership cost per vehicle. Machine learning techniques outperformed the linear regression method in terms of predictive power and robustness (understood as the dispersion of predictive power measures over the training sets). Based on SHAP values, the sensitivity analyses showed that the proposed models could be used to predict the impact of changes in some critical features on corresponding unit costs. The results may be useful for: (i) introducing regulatory constraints to the allocation of national public resources to local public bus transport services, aimed at minimizing the resources needed to provide a given level of service; (ii) defining the maximum economic compensation required by firms involved in competitive tendering for the allotment of service concessions, or firms with monopoly rights (by political choice and/or local public ownership); and (iii) improving service contract management and design by identifying key cost drivers of transit services.

Keywords

Local public transport; cost models; machine learning; transportation policy.

1. Introduction

In the United States and Europe, national and local governments play an active role in public transportation by providing subsidies to local public transport (LPT) service operators and funding expenditures (Verdeaux, 2003). A central issue in reform initiatives is the design of reimbursement systems, which usually reward operators for providing LPT services at cost-efficient levels.[1] For this reason, it is crucial that policymakers in the LPT industry have access to analytical tools for cost prediction.

The present research aimed at developing several machine learning (ML)-based cost models to predict an efficient total economic cost per vehicle revenue-mile of local public bus transport (LPBT) services. Accurate cost forecasting is essential for making informed decisions about budget allocations and resource utilization (Chou, 2009), as well as to improve service contract management and design. First, by exploiting the favorable incentive properties of yardstick competition (Shleifer 1985), a forecasting model might be used at a micro-level to define the maximum economic compensation required by firms involved in competitive tendering procedures for the allotment of service concessions, or firms with monopoly rights (by political choice and/or local public ownership). Second, at a macro-level, a forecasting model might be employed by policymakers to introduce regulatory constraints to the allocation of national public resources earmarked for LPBT services, aimed at minimizing the total resources needed for a given level of service. Third, if empowered with specific features (e.g., scale of service), a forecasting model might provide cost estimates that convey critical information for the design of (competitive or uncompetitive) procedures for the allotment of LPBT services (with respect to, e.g., the determination of network size or the number of vehicle revenue-miles). Fourth, a forecasting model might identify key cost drivers, enabling policymakers to predict the impact of changes in some critical features on corresponding unit costs. This enables more efficient management of service contracts by providing information on key variables to be monitored and negotiated with transportation operators, and thus improving transparency in the use of public subsidies.

The present study aimed to develop a methodological framework for incorporating ML approaches into the prediction of the total economic cost per vehicle revenue mile (vrm)

[1] See, e.g., Norway (Dalen and Gomez-Lobo, 1996, 2003), France (Gnagnepain and Ivaldi, 2002; Roy and Yvrande-Billon, 2007), Italy (Avenali et al., 2018)).

of LPBT services. Six ML methods were considered and compared with the conventional approach of *Multivariate Linear Regression*: *K-Nearest Neighbors*, *Support Vector Regression, Multilayer Perceptrons, Random Forest, Gradient Boosting*, and *XGBoost.*. The outcome is a set of ML-based cost models, where the predicted variable is the total economic cost per vrm of LPBT services and the predictors are features selected case by case, depending on (i) the nature of the data and (ii) the tradeoff between predictive power (accuracy and robustness) and interpretability.

Despite the significant predictive capabilities of ML methods, their "black box" nature often limits interpretability. To address this challenge, a SHAP analysis was implemented (Lundberg and Lee, 2017), providing a deeper understanding of each feature’s contribution to the model’s outcome, namely in relation to cost predictions.[2] The utilization of SHAP plots, in conjunction with policy goals, allows for further sensitivity analyses to assess the impact of relevant features on unit costs. Thus, sensitivity analyses were applied to highlight the marginal impact of efficiency gains obtained by manipulating cost-driving variables (controlled by either transport operators or public authorities). In particular, the present study considered: (i) features related to the service context (e.g., commercial speed), to consider the potential impact of congestion and orography on costs; (ii) features related to the service size (e.g., network size, vehicle revenue-miles), to highlight density and economies or diseconomies of scale; and (iii) features related to the production process (e.g., vehicle productivity, average hourly wage, percentage of electric vehicles).

To train the models, data were collected from 269 transit agencies providing urban services in the United States between 2015–2019 (representing 85% of all agencies operating urban bus services in the United States). The primary data sources were the National Transit Database of the Federal Transit Administration (FTA) and the Public Transportation Vehicle Database of the American Public Transportation Association

[2] Usually machine learning models are black boxes where the interpretation of the features roles is difficult. SHAP values (SHapley Additive exPlanations) are a technique which relies on the Shapley values proposed within cooperative game theory (Shapley, 1951, 1953), and indeed estimates the marginal contribution of each feature by generating several permutations of the features and then combining the results obtained in terms of prediction differences from a reference prediction.

(APTA). These are public repositories of data on the financial, operating, and asset conditions of American transit systems.

The paper is organized as follows. Section 2 comprises a review of the relevant literature. Section 3 presents the dataset and methodological framework of the present study. Section 4 presents model selection and results. Section 5 provides the results and policy implications of the present work, and Section 6 concludes the discussion.

2. Literature review

The relevant literature can be broadly divided into two major streams: (i) studies examining the cost structure of LPBT companies and LPBT cost forecasting, and (ii) studies examining ML applications for transport policy and cost estimation.

Local public bus transport cost structures and forecasting

Studies exploring the cost structures of LPBT companies have mainly focused on the estimation of variable and total costs. However, there has been significant debate over the determination of input and output measures (see Daraio et al., 2016, for a detailed survey). Input variables normally fall into two main categories: (i) "physical" production factors with their own measurement units (e.g., number of employees, number of driving vs. non-driving employees, work hours, fuel consumption, number of vehicles in the fleet); and (ii) monetary costs, split into capital expenses (CAPEX) and operating expenses (OPEX). On the output side, variables used to measure production efficiency include vehicle-kilometers (e.g., Cambini and Filippini, 2003), seat-kilometers (e.g., Farsi et al., 2007), and total-seat-kilometers (e.g., Gagnepain and Ivaldi, 2002). All of these measures are from the supply side. In contrast, variables related to production effectiveness (e.g., number of passengers, passengers by kilometer travelled) fall on the demand side (e.g., Bhattacharyya et al., 1995). Hedonic characteristics include service frequency, average commercial speed, and average fleet age (e.g., among others, Fraquelli et al., 2004; Shaw et al., 2005; Piacenza 2006; Cambini et al., 2007).[3] Additionally, there is a lack of

[3] Higher speed implies lower operating costs, but also better service (i.e., passengers are transported in less time). Commercial speed also relates to external factors that are often outside operators' control (i.e., the presence of preferential lanes or other measures to reduce congestion). Average fleet age refers to both efficiency and effectiveness, since a younger fleet is usually less expensive (in terms of fuel consumption and maintenance costs), has higher perceived quality (according to passengers), and emits less pollutants.

consensus on LPBT operators' economies of scale. Boitani et al.'s (2013) cross-country analysis of 77 LPT companies operating in large European cities found diseconomies of scale, while Fraquelli et al. (2001) found that the average cost per seat-kilometer was U-shaped, consistent with Avenali et al. (2016). Other studies have found economies of scale for urban systems, suggesting that such systems would not recover costs through marginal cost pricing (e.g., Cambini et al., 2007; Farsi et al., 2007). Specifically, Viton (1992), Colburn and Talley (1992), and Ripplinger and Bitzan (2018) investigated the cost structure of transit agencies in urban communities in the United States (the focus of the present study). The results indicated that the firms exhibited economies of scale over a wide service range.

LPBT service cost structures may be summarized as follows: (i) they are labor (rather than capital) intensive, with the cost of fuel and ownership representing other important dimensions (see Table 3, which presents detailed information on the transit agencies included in the present sample); (ii) among the hedonic characteristics, commercial speed mainly explains their cost differentials, while average fleet age also plays a role; and (iii) there is no consensus on the presence of economies of scale in the provision of services.

With regard to cost forecasting, Hensher et al. (2013) introduced a simplified performance-linked payment (SPLP) model that could be used to assess public subsidies for LPBT operators. The model internalized the effects of exogenous variables (not under the operator's control, such as commercial speed) on the cost of LPBT services. However, the parameter estimates were not representative of any specific operating context, but based on reasonable assumptions for Australian metropolitan areas. Avenali et al. (2016, 2018) estimated the unit cost of LPBT services in Italy using a piecewise regression model; they found that commercial speed was the most important cost driver, while economies of scale were low and limited to only small service bunches. The results also highlighted a positive correlation between bus fleet investments and the total service cost.

In this context, the contribution of this paper lies in two main aspects. First, accuracy and robustness of the cost predictions of six ML methods – i.e., K-Nearest Neighbors, Support Vector Regression, Multilayer Perceptrons, Random Forest, Gradient Boosting, and XGBoost – are compared with those of linear regression (i.e., a traditional parametric approach). This allow us to assess significant improvements in public funding allocation enabled by a higher predictive power of ML approaches. Moreover, all of these methods were compared both with and without feature selection, resulting also in identifying most influential cost drivers for urban bus transit service. Second, leveraging an integrated database (covering more than two-thirds of LPBT service production in the United States and nearly 85% of passengers carried from 2015–2019), the methodological framework incorporated ML approaches for the purpose of supporting policymaking in the transport sector (rather than supporting political ends), with respect to a particular orography (i.e., that of the United States).

Machine learning applications for transportation policymaking and cost estimation

ML approaches are becoming increasingly applied in the field of transportation policy and practice (Tizghadam et al., 2019). While some research has discussed how ML methods may be utilized to improve the performance of transportation data analytics tools (focusing on the quality and quantity of available data; e.g., Bhavsar et al., 2017), most studies have aimed at developing ML approaches to predict transportation system dimensions, including transportation demand (Salas et al., 2022; Plakandaras et al., 2019; Hagenauer and Helbich 2017), rail network performance (Gunduz et al., 2011; Li et al., 2014), traffic conditions (Jacob and Abdulhai, 2010; Liu et al., 2019, Ma et al., 2020; Yang et al. 2020; Raju et al., 2022), road maintenance (Mahpour and El-Diraby, 2022), aircraft boarding (Schultz and Reitmann, 2019), bike-sharing demand (Xu et al., 2018; Gammelli et al., 2020) and inventory (Ceccarelli et al., 2021), dial-a-ride system planning (Marković et al., 2016), and electric vehicle performance (Chen et al., 2021; Liu et al., 2021). While ML approaches to cost estimation are widespread in the literature, no prior research has developed an ML approach to predict the cost of transportation services. For instance, studies have proposed ML models for the purpose of estimating software costs (e.g., Huang et al., 2015; Catal, 2011), memory systems (Servadei et al., 2020), product life cycles (e.g., Yeh and Deng, 2012), construction projects (e.g., Hashemi et al., 2020), supply chain components (Bodendorf et al., 2021), and customized furniture manufacturing (Kurasova et al., 2021), among other components.

3. Methodology

This section describes the methodology that was applied in the present study to predict the unit costs of LPBT services and to identify key cost drivers. Figure 1 displays the analytical framework, which involved three main steps: (1) data collection (Section 3.1), (2) data pre-processing (Section 3.2), and (3) the implementation of predictive methods (Section 3.3, which also summarizes the main features of the six supervised ML models).

3.1 Data collection

The FTA National Transit Database – a public repository of data on the financial, operating, and asset conditions of American transit systems – was the primary source of the analyzed data. Multiple NTD reports and annual data tables were then integrated into the FTA data. Specifically, for each year, relevant NTD data were taken from the "Service Annual Data Tables," "Fuel and Energy Annual Data Tables," "Employees Annual Data Service," "Annual Database Operating Expenses," and "Annual Database Revenue Vehicle Inventory." In the Annual Database Revenue Vehicle Inventory, each item describes a specific purchase batch, including detailed information on fleet characteristics (e.g., number of buses, manufacturing year, vehicle size, power/fuel type, seating and standing capacity, useful life benchmark, average lifetime miles, annual miles).

The NTD data lacked some critical information for the present analysis (i.e., bus purchase price), which was subsequently sourced from the APTA Public Transportation Vehicle Database. This enabled the analysis to estimate two crucial variables: annual vehicle depreciation and the residual value of vehicles. Following the literature (Williams, 1979; Viton, 1981; Berechman, 1987; Filippini and Prioni, 1994; Karlaftis and McCarthy, 1999; Piacenza, 2006; Cambini et al., 2007), the number of vehicles owned by each transit agency was used as a physical proxy of invested capital, while the net value of those vehicles was considered a proxy of the net invested capital. Therefore, the cost of capital was estimated by multiplying the residual value of the bus fleet with the weighted average cost of capital (WACC). In this case, the after-tax WACC was set to 4.41%, in accordance with Damodaran's (2021) estimate for the United States transportation sector.

The result was a built-in database of 269 full reporter transit agencies. For each agency, disaggregated data were available, with reference to general information, service features, personnel management, fleet characteristics, and economic costs. Monetary values were in 2019 prices (USD) and inflated in line with the United States Consumer Price Index. Table 1 presents the database details.

[Table 1 near here]

The transport operators included in the sample provided more than 1.4 billion vehicle revenue-miles, representing more than two-thirds of all LPBT service production in the United States and over 80% of all passengers carried. Of note, the dataset was characterized by a high level of inter-agency variability (see Tables A.1, A.2, and A.3 in

Appendix A for descriptive statistics). Specifically, the sample comprised transit agencies producing less than 100,000 vehicle revenue-miles per year and others offering more than tens of millions; transit agencies with an average fleet age of less than 3 years and others with a fleet aged more than 12 years; transit agencies with minibuses with fewer than 20 seats and others with articulated buses with a capacity of more than 80 passengers; transit agencies that relied on conventional fuel buses and others that were early adopters of alternative power technologies (e.g., hybrid, full electric); transit agencies operating in large, congested cities (e.g., average speed lower than 10 mi/h) and others offering transit services in towns where traffic moved smoothly; and transit agencies taking different approaches to personnel management and employee remuneration, also influenced by local/state legislation and regulation.

3.2 Data pre-processing

To ensure that accurate and meaningful insights could be extracted from the data, a series of data pre-processing operations were employed.

Feature engineering was implemented to build a dataset suitable for ML. In particular, important features (e.g., network turnover, percentage of transport operators, operator productivity, fleet ownership cost) were defined, combining individual components in Table 1, to constitute the final set of 39 features (see Table A.4 in Appendix A). The relevant input variables were identified according to the empirical and theoretical literature in the urban public transport sector (Daraio et al., 2016).

The definition of *total economic cost* was based on the expense categories described in Table 2. In line with the literature (e.g., Wunsch, 1996; Avenali et al., 2018), the total economic cost per vehicle revenue-mile (i.e., the ratio between the total economic cost and the service size, measured as the number of vehicle revenue-miles) was estimated (henceforth referred to as the *total cost per vrm*, or *unit cost*). The total economic cost of the transit service was calculated as the sum of the cost components presented in Table 3, which also presents the average share of different cost items. Vehicle operating costs accounted for 46% of the total cost per vehicle mile, with more than 67% of these costs representing transport operators' salaries and wages. Efficiency in personnel management (e.g., transport operators' productivity or average hourly wage) was expected to be an important predictor of average unit cost, as expenses connected to fleet management significantly affect the economic performance of transit agencies. Indeed, almost 30% of the total costs of these agencies relate to vehicle maintenance and depreciation. Consequently, vehicle usage (i.e., productivity, in miles) and characteristics (e.g., average age, seat and standing capacity) are critical factors for transit agency efficiency.

[Table 2 near here]

[Table 3 near here]

The *annual dataset integration* consisted of merging all potential predictors and their unit costs into a single multiannual dataset. At this point, the data were explored and the quality of the dataset was verified (according to the absence of errors and duplicate data). Due to the differing scale of features, data normalization was applied. This resulted in a panel database of 1,345 observations related to 269 United States transit agencies that directly provided LPBT services from 2015–2019. To reduce data noise (e.g., smooth outliers) (Zimek et al., 2013), the features and output values (i.e., total cost per vrm) were averaged over the five early datasets from 2015–2019, to obtain the final dataset.

Finally, a *hold-out set* strategy was applied. To assess the generalization of a predicting model, the dataset was split into a training set (T_r) (as in Ban et al., 2013) (used in the training phase) and a testing set (T_s). The ability to accurately predict the output of unseen observations depends, in part, on different extractions of T_r and T_s. K-fold cross-validation was applied to evaluate the final performance of the model, with the average performance obtained over all k folds (i.e., different extractions) (Hastie et al., 2009). The dataset was split by randomly extracting a test set T_s, comprised of 20% of the data, and a training set T_r, comprised of the remaining 80%. To obtain reliable and balanced results, the sample was divided into groups based on quartiles of the output values, and an equal portion was extracted from different sample subsets, defined on the basis of the output value quartiles. Moreover, five random T_s extractions were performed to determine five different pairs of T_r and T_s (henceforth referred to as *extractions* and denoted 0–4), and all training and testing analyses were conducted for each pair. Thus, for each extraction, there were 215 T_r data and 54 T_s data.

3.3 Predictive method

This section describes the principles and application of the six ML methods that were applied in the present research to identify possible complex non-linear relationships between LPBT unit costs and service features.

3.3.1 Principles of the predictive methods

The present study considered six ML techniques, comparing their performance to that of the traditional parametric approach of linear regression. In the Multivariate Linear Regression (MLR) parametric model (see, e.g., James et al., 2013), the input-output process is approximated as a linear function, with the output computed as a linear combination of features (also denoted *predictors*). As MLR is easy to compute and highly interpretable, it has been widely applied in transportation economics (e.g., Boitani et al., 2013; Ripplinger and Bitzan, 2018). However, its predictive power may suffer when the process is characterized by non-linear interactions between the input and the output, as is often the case.

K-Nearest Neighbors (KNN) regression is one of the simplest non-parametric models. For a given value k (set by the user), it predicts the output $\hat{y}$ in correspondence of an input vector x, as the average output of the set of k samples in T_r that are closest to x (Bishop, 2006). Due to its simplicity, the KNN method is often applied in practice (see, e.g., Ban et al., 2013; Kohli et al., 2021). However, it requires a large amount of memory for training, especially with a larger number of features, and its distance-based mechanism makes it sensitive to outliers and data imbalance.

Support Vector Regression (SVR) is the simple regression adaptation of the Support Vector Machine (SVM) classification method (Boser et al., 1992). SVM training is formulated as a convex-constrained optimization problem; in principle, this can be solved efficiently. However, the dimension of the training problem rapidly increases in line with T_r size. Nonetheless, many decomposition algorithms have been applied to big data in a reasonable amount of time (see, e.g., Fan et al., 2008; Chang and Lin, 2011; Manno et al., 2016; Manno et al., 2018), making the method very appealing for both classification and regression.

Multilayer Perceptron (MLP) represents one of the most widely used neural network (NN) architectures (see, e.g., Bishop, 1995) in supervised ML. NNs are input-output structures composed of many processing units (i.e., neurons) that are connected to each other by weighted oriented connections. Each neuron elaborates the weighted sum of signals received from its incoming connections by means of a non-linear *activation* function to produce an outgoing signal, which is subsequently propagated to adjacent neurons. In MLP, neurons are organized into layers and connections are oriented from the input to the output layer (see, e.g., Grippo et al., 2015). The MLP structure is particularly well suited to capturing complex non-linear relationships between variables. Moreover its universal approximation property has been demonstrated (see Leshno et al., 1993). Accordingly, the MLP approach has been extensively applied in a variety of contexts (see, e.g., Qasim and Khadkikar, 2014; Laboissiere et al., 2015; Chelazzi et al., 2021; Manno et al., 2021; Manno et al., 2022). Nonetheless, the highly non-linear and non-convex optimization problem that arises in the training of MLP remains a significant challenge.

Random Forest (RF) (Ho, 1995) is a classification and regression ensemble method (see Dietterich, 2000) that is trained by building many decision trees (DTs) (Breiman et al., 2017). In the regression task, the final prediction is based on an average of the DT predictions. This is done to prevent overfitting, which often occurs when a single DT is used.

Gradient Boosting (GB) (Friedman, 2001) is a similar ensemble technique in which new DTs are iteratively added to reduce the errors made by already inserted trees. The gradient descent algorithm (see, e.g., Bertsekas, 1999) is used to minimize the loss function when adding a DT, and the iterative addition stops when no significant improvements are obtained.

XGBoost is an efficient implementation of GB that has successfully addressed many ML challenges (Chen and Guestrin, 2016). For this reason, it has recently been applied to solving many practical problems (e.g., Pan, 2018; Li et al., 2019; Zheng and Wu, 2019).

The above ML techniques reflect three different categories of approaches, which can be compared with the more conventional MLR approach. KNN is very simple and easy to use (thanks mainly to the straightforward training and the presence of only one hyperparameter to be tuned), but sometimes not sufficiently accurate. MPL and SVR are well suited to capturing non-linear relationships, so tend to be very accurate; however, they are less interpretable and prone to overfitting. RF, GB, and XGBoost, combining weaker and more interpretable learners, aim at a trade-off between predictive power and interpretability. Nevertheless, SHAP analysis was implemented to provide a deeper understanding of each feature’s contribution to the model outcomes, also gaining insights into the factors that influence cost predictions.

3.3.2 Description of the predictive experiments

The described methods were applied to predict the *total cost per vrm*, which represents the dependent variable in all the models. MLR, KNN, SVR, MLP, RF, and GB were implemented using the Python sklearn package, while XGBoost was taken from https://xgboost.readthedocs.io/en/stable.[4]

The independent variables (i.e., the potential predictors) for the unit cost estimation given as input to each model are the result of the feature engineering stage described in Section 3.2. The MLR and six supervised ML models were compared under two different conditions:

a. without features selection (i.e., considering all 39 features), and
b. with feature selection.

Feature selection can be used to improve model interpretability, applicability, and possibly predictive power (see, e.g., James et al., 2013). Commonly, some or many of the variables used in regression and classification models are not significantly relevant to/for the output, and they may generate unnecessary complexity in the resulting model. The removal of these variables can result in a model that is more easily applicable and interpretable. Moreover, their removal also simplifies the training optimization problem, which may substantially improve predictive power.

Concerning the experiments without feature selection, the independent variables comprise all 39 features included in our dataset (Table A.4). In this case, for each of the five extractions and for each ML method, a grid search with 5-fold cross-validation was applied to the T_r to determine the best hyperparameter values. Subsequently, each ML method was trained with these hyperparameters on the entire T_r and tested on the T_S. Table A.5 (Appendix A) reports the hyperparameter values used in the grid search.

With regard to the feature selection experiments, three different feature selection methods were considered: forward stepwise selection (FSS), backward stepwise selection (BSS), and recursive feature elimination (RFE). Briefly, FSS starts with an empty model and iteratively adds the most relevant features, BSS starts with a model with a full set of features and iteratively eliminates the most irrelevant ones, and RFE exploits the weights

[4] All experiments were conducted on a laptop AMD Ryzen 5 3500U 2.10 GHz with 8 GB of RAM.

assigned to features by the ML method to recursively select smaller and smaller sets of features. The present analysis focused on the results obtained using BSS, as this emerged as the best performing method in terms of the proportion of variance explained. The feature selection experiments were repeated for each extraction, method, and maximum number of selected features (ranging from 4–11). Moreover, for each method, feature selection was applied using the hyperparameter values obtained in the grid search of the prior experiment. Once the subset of features was determined, a further grid search using 5-fold cross-validation was applied to determine new hyperparameter values that were better suited to the restricted setting. Subsequently, these hyperparameter values were used to train the method on the restricted T_r and test the trained model on the T_s.

Of note, in the preliminary experiments, a "manual" feature selection method was tested, whereby multiple subsets of a few critical features were selected according to the degree to which they were assumed to explain unit costs and their ease of finding. Since the manual feature selection experiments yielded substantially lower accuracies than the standard feature selection approaches, they are not reported in the following.[5]

For each method, extraction, and maximum number of selected features, the feature selection experiment that achieved the best performance on the corresponding T_s is reported. The coefficient of determination (R^2), understood as a measure of model accuracy (i.e., predictive power), was used as the performance metric. Below, this method is referred to as the "*best test set*" approach.

4. Model selection and results

Table 4 displays the outcomes for the seven methods, both with and without feature selection, for each training set. It also reports, for each method: (i) the average R^2 score (i.e., a measure of predictive power) over the five training sets and (ii) the standard deviation of R^2 (i.e., a measure of robustness) over the five training sets.

[Table 4 near here]

First of all, concerning the experiments without feature selection, all the supervised ML models (except KNN) outperformed the standard MLR in terms of accuracy. In particular,

[5] The parsimonious paradigm, whereby ML techniques automatically select the best trade-off between predictive power and model simplicity (i.e., number of features), was also run. However, the main findings essentially coincided with the presented results.

MLR achieved a mean R2 testing accuracy of 0.748, RF obtained 0.782, and the mean R2 of SVR, MLP, GB, and XGBoost were all above 0.8, with MLP resulting the best performing one with R2 = 0.868. The poorest performance was obtained by the KNN model which showed to be too simple for the considered dataset.

As is easily observed in Table 4, the feature selection strategy was very effective for each method, in terms of both predictive power and robustness. The only exception was related to the robustness of MLR (0.043), which outperformed that of MLR with feature selection (0.062). This was due to an improvement in the R^2 score, which was more significant for some training tests over others. In general, feature selection improved predictive power and robustness while simultaneously simplifying the models (which selected, on average, only one-fourth of the features). Even when feature selection is applied, four supervised ML models (i.e., SVR, MLP, GB, and XGBoost) outperform the MLR method in terms of both the average amount of explained variance and robustness with respect to the different training sets.

The research aimed at identifying models that would provide a good trade-off between model complexity and predictive power. Therefore, particular attention was given to models run under the best test set approach. Specifically, to maximize predictive power, for each training set 0, 1, 2, 3, and 4, the best model (in terms of R^2) was selected using the best test set approach; these were, respectively, SVR (R^2 = 0.846, 10 features), GB (R^2 = 0.936, 11 features), MLP (R^2 = 0.927, 7 features), MLP (R^2 = 0.918, 9 features), and SVR (R^2 = 0.924, 10 features).

These models extracted information from a subset of the available features. However, some interesting and widely available features were never exploited for the estimation of unit cost, even if they were incorporated into the highly predictive models. To mitigate this drawback, one further model was selected that included the additional feature of *network turnover*, defined as the ratio between service scale (in vehicle revenue-miles) and network size (in directional route miles).[6] This feature provides critical information for the design of (competitive or non-competitive) procedures for allotting LPBT services. The model with the largest R^2 that also incorporated network turnover as a

[6] Directional route miles are the total number of miles in each direction that public transportation vehicles travel during revenue service, measured to the nearest hundredth of a mile and specified for each combination of mode and service with a fixed guideway.

feature was provided by the MLP method, constrained to 11 features under training set 3 ($R^2 = 0.904$).

Table 5 describes the six models applied to predict the unit cost of any LPBT service using the selected features. All models are labeled as "*predictive method* (MLR, KNN, SVR, MLP, RF, GB, and XGBoost) – *extraction* ('e') of the training set (0, 1, 2, 3, 4) – number of *selected features* ('f')". For example, SVR-e0-f10 represents the model based on the Support Vector Regression (SVR) method, trained on extraction 0 of the training set, and using 10 features selected by the backward stepwise selection (BSS) method.

As can be observed, given the objective of maximizing the accuracy of the cost prediction, no MLR model was selected, as there was always a more effective model.

[Table 5 near here]

To gain a deeper understanding of the underlying mechanisms driving model prediction, the SHAP values of every feature included in the proposed models were plotted (see Appendix B). This enabled the contribution of individual features to be identified, in terms of the direction and scale of their impact on the model outcome. For each proposed model, only a subset of the features determined larger impacts on unit costs. In addition, in several cases, there was overlap in the high-impact features between models (with respect to, e.g., average salary expenses per employee, operator productivity (annual miles), average speed, vehicle productivity (annual miles), and fleet ownership cost per vehicle).

The proposed models were tested for a set of six LPBT services, of which three were provided by transit agencies that were not present in the database (i.e., from Lancaster, Sheboygan, and Tucson). Table 6 reports the features related to the models in Table 5 for these six services and the unit costs and average values predicted by each model.

[Table 6 near here]

The results demonstrated a high degree of predictive power in forecasting the total cost per vrm in the six case studies, with an average percentage error of approximately 6% between the predicted and the observed values. Moreover, the models showed low variance with respect to the different training sets, indicating robustness and reliability across a range of input data.

Since the proposed models were trained on real-world LPBT services, they may be used by public authorities to define an upper bound for the economic compensation of firms (both in competitive tendering procedures and when LPBT service allotments are not tendered out), so as to exploit the incentive properties of yardstick competition (Shleifer, 1985). Indeed, the models' predicted unit costs represent benchmark figures, achieved by an implicit comparison of the unit costs of real LPBT services (as the benchmark figures were generated by techniques trained on real-world data, they represent, for any bunch of identical or very similar services, "average" values of the corresponding unit costs).

Within this framework, policymakers can select the ML models most suitable for their specific case based on the available data in their context. Once the models are developed, validated, and tailored to a specific country, public authorities can collect the necessary inputs for the proposed models. These inputs are generally either readily available, as they pertain to the operational characteristics of the LPBT service (e.g., service miles traveled, average commercial speed, fleet characteristics, route network extensions, cost of fleet ownership per vehicle), or are assumed to be target values determined by policymakers (e.g., operator productivity in terms of annual miles, average wage expenditures per employee).

5. SHAP-based sensitivity analysis

Public authorities may use the proposed models to estimate the impact of some variations in specific service features on expected unit costs. In particular, Figure 2 shows the impact of a subset of "core" features identified by the SHAP plot analysis of the proposed models, namely: average speed, average salary expenses per employee, vehicle productivity (annual miles), and fleet ownership cost (depreciation and cost of invested capital) per vehicle.[7]

[Figure 2 near here]

The proposed models highlighted that unit cost decreased as commercial speed increased. Indeed, lower commercial speed typically reduces the annual productivity of driving personnel. Thus, LPBT service provision may require a higher number of drivers (usually, the shift duration of any driver cannot extend over a pre-defined upper bound, also for safety reasons). In addition, higher commercial speed was associated with less fuel

[7] Figures 2 and 3 report the average values of all six predictions of the applied models.

consumption and less vehicle stress (by minimizing stop-and-go driving, which can negatively affect the engine and transmission system and increase maintenance costs).

As expected, unit costs tended to increase in line with the average employee salaries, highlighting the importance of the local context in determining LPBT costs. This effect was particularly evident in larger cities, where wages were typically higher, generating higher total cost per vrm (see, e.g., the case of Washington). Additionally, lower annual vehicle productivity was associated with higher unit costs, as more vehicles were required to produce the same amount of vehicle revenue-miles. This finding has implications for cases when more service is needed during on-peak periods, but no further vehicles are available and thus new ones must be purchased/rented. Finally, bus fleets with higher unit ownership costs induced larger unit costs. This result was strictly linked to vehicle purchase costs and the cost of capital for transport providers.

Even though the network turnover (and directional route miles) and the percentage of electric vehicles in the fleet were not highlighted as “core” features by the SHAP plot analysis, the effects of their possible changes were investigated since these features are of growing importance to policymaking.

[Figure 3 near here]

The proposed models highlighted that a raise in network turnover (given a fixed network and route structure) could generate a density economy and thus reduce the expected unit cost (see Figure 4); on the other hand, within congested metropolitan areas (see, in particular, the case of Washington), greater network turnover could generate a density diseconomy (given a fixed network and route structure), thereby raising the expected unit cost. Additionally, in the present analysis, as directional route miles increased while network turnover remained constant, possible scale economies could lower the predicted unit cost.

Finally, Figure 3 shows that fleets with a higher percentage of electric vehicles were associated with larger unit costs, mainly due to the much more expensive depreciation and cost of the net invested capital (consistent with the literature on public transport electrification, e.g., Comello et al., 2021). Interestingly, as fleet sizes became very large (see, in particular, the case of Washington), unit costs increased at a decreasingly marginal rate.

The results of the sensitivity analyses were consistent with the initial expectations. This is particularly interesting, given that ML models incorporate non-linear functions, and it cannot be assumed that variations in key cost drivers will necessarily generate a coherent impact on the output. The present findings suggest that the ML models accurately captured the complex relationships between service features and unit costs in LPBT services, thereby increasing confidence in the validity of their predictions and their utility for policymakers.

6. Conclusions

The present study aimed to develop a methodological framework for incorporating ML approaches into the prediction of the total economic cost per vehicle revenue mile (vrm) of LPBT services. Six ML methods were considered and compared with the conventional approach of Multivariate Linear Regression. The outcome is a set of ML-based cost models, where the predicted variable is the total economic cost per vrm of LPBT services and the predictors are features selected case by case, depending on (i) the nature of the data and (ii) the tradeoff between predictive power (accuracy and robustness) and interpretability. The developed framework was implemented on a dataset containing information from 269 transit agencies providing urban services in the United States from 2015 to 2019.

When comparing standard MLR against six ML models on the dataset including all available features, five out of six ML models outperformed MLR in terms of testing accuracy, with MLP emerging as the most preferable one. Further results showed that the feature selection strategy was very effective for each method, in terms of both predictive power and robustness (understood as the dispersion of predictive power measures over the training sets). Moreover, the application of feature selection had the benefit of providing simpler models (incorporating, on average, only one-fourth of the available features). Even when feature selection is applied, four supervised ML models (i.e., SVR, MLP, GB, and XGBoost) outperform the MLR method in terms of both the average amount of explained variance and robustness with respect to the different training sets. This means that ML models have proven to be a more reliable approach with respect to different input data available for transport planners.

Our application demonstrates the following key points: (i) ML models outperformed MLR in both testing accuracy and robustness. Robustness, measured as the dispersion of predictive power across training sets, indicates that ML models provide a more reliable approach for transport planners given varying input data; (ii) ML models, such as MLPs and SVR, are adept at capturing non-linear relationships and tend to be highly accurate. Our dataset encompasses a diverse range of services, differing in size and service context (e.g., variations in rolling stock, percentage of low-emission vehicles, and average bus age). Traditional methods, such as linear regression, struggle to incorporate these complex non-linear relationships. In contrast, MLPs and SVR excel in modeling these relationships due to their superior expressive power (Bishop, 2006).

Then, the primary advantage of using ML methods lies in their improved accuracy and robustness in cost prediction, particularly in the presence of non-linear relationships within heterogeneous datasets. Accurate cost forecasting is crucial for informed decision-making regarding budget allocations and resource utilization (Chou, 2009), as well as enhancing service contract management and design.

Moreover, the integration of SHAP analysis enhanced the interpretability of the ML models, facilitating more informed decision-making within the scope of this research. Indeed, for each proposed model, a subset of features determined a large impact on unit cost. These “core” features included: average salary expenses per employee, average speed, vehicle productivity, and fleet ownership cost per vehicle. The sensitivity analyses showed that unit cost decreased as commercial speed increased, and unit cost increased as annual vehicle productivity decreased, as more vehicles were required to produce the same number of vehicle revenue-miles. Additionally, costs tended to increase in line with the average employee salary and the unit ownership cost associated with the fleet. Therefore, the results of the sensitivity analyses suggest that transport planners can exploit the proposed models to accurately predict the impact of changes in some critical features on corresponding unit costs.

Regarding some relevant policy variables, in some cases, increased network turnover was found to generate a density economy and thus reduce expected unit costs; on the other hand, in congested (metropolitan) areas (e.g., Washington), increased network turnover was found to generate a density diseconomy, thereby increasing the expected unit costs. Interestingly, fleets with a higher percentage of electric vehicles were associated with larger unit costs, mainly due to the much more expensive costs of depreciation and net invested capital. However, as fleet size became very large (see, e.g., the case of Washington), unit costs increased at a decreasingly marginal rate.

The present study suffered from some limitations, which may indicate directions for future research. First, as the models were top-down, they enabled prediction but were "black boxes" for transit operators. A bottom-up approach would both predict (average/maximum) efficient unit costs and make inefficiency causes clear to operators, highlighting key competitive insights. Second, when feature selection was applied, the ML methods focused on a restricted subset of features that were identified from a starting set of 39 features. Future research could re-run the analysis while constraining all methods (i.e., with and without feature selection) to learn from only a few critical features, such as the most frequently available features and the features whose measures relate to actual characteristics of the required service.

Tables

Table 1 - Structure and individual components of the built-in database

Major category	*Individual component*
General information	Agency ID
	Reference year
	State
	City served
	Population of primary urbanized area served
Service features	Vehicle-miles
	Vehicle-revenue miles
	Directional route miles
	Unlinked passenger trips
	Passenger miles
	Average speed
	Fuel used (gallon equivalent)
Personnel management	Number of total employees
	Percentage of full-time employees
	Average salary expenses per employee
	Number of transport operators (i.e., drivers and movement personnel)
	Number of hours worked by transport operators
	Average hourly wage of transport operators
Fleet characteristics	Number of active fleet vehicles
	Average fleet age
	Average fleet length
	Average fleet capacity (seats + standing)
	Vehicles productivity (annual miles)
	Average lifetime miles of active fleet
	Percentage of power/fuel types
Economic costs	Operating expenses related to vehicle operations
	Operating expenses related to transport operators wages
	Operating expenses related to fuel/energy and lubricants
	Operating expenses related to vehicle maintenance
	Operating expenses related to facility maintenance
	Operating expenses related to general administration
	Vehicles depreciation costs
	Cost of invested capital

Table 2 - The categories of expenses included in the cost basis

Category of expenses	Cost items
OPEX related to vehicle operations	Expenses related to activities associated with dispatching and running vehicles to carry passengers. We define four main sets of items: - OPEX related to transport operators (i.e., drivers and movement personnel) wages; - OPEX related to fuel/energy and lubricants; - Vehicle depreciation costs - OPEX related to other vehicle operation activities (e.g., tires and tubes, outsourcing services and miscellaneous).
OPEX related to vehicle maintenance	Expenses incurred during all activities related to keeping vehicles operational and in good repair (e.g., maintenance workers' salaries, spare parts, outsourcing maintenance, utilities and miscellaneous).
OPEX related to facility maintenance	Expenses include all activities related to keeping depots, structures, roadways, and other non-vehicle assets operational and in good repair (e.g., facility maintainers' salaries, materials and supplies).
OPEX related to general administration	Expenses incurred to perform support and administrative activities (e.g., overall management, economic planning and control costs, membership fees, business consulting and information systems costs, salaries of personnel employed in general activities).
Cost of invested capital	The cost of a transit agency's funds (debt and equity) with regard to urban bus services.

Table 3 – The components of the average unit cost related to LPBT services

Cost per vehicle-mile (USD/vrm)	*%*	*Mean*	*Min*	*1° quartile*	*Median*	*3° quartile*	*Max*	*Coeff. of variation*
OPEX related to vehicle operations	45.8%	3.65	1.10	2.81	3.34	4.22	12.33	0.37
OPEX related to transport operators wages	67.8%	2.47	0.33	1.83	2.30	2.91	9.46	0.41
OPEX related to fuel/energy and lubricants	15.0%	0.55	0.18	0.42	0.51	0.64	1.40	0.33
OPEX related to other vehicle operation activities	17.2%	0.63	0.02	0.39	0.55	0.77	3.03	0.60
OPEX related to vehicle maintenance	15.1%	1.21	0.18	0.83	1.10	1.40	5.60	0.50
OPEX related to facility maintenance	3.5%	0.28	0.00	0.13	0.22	0.36	3.47	0.91
OPEX related to general administration	17.7%	1.41	0.15	0.89	1.22	1.70	7.92	0.59
Vehicles depreciation costs	14.2%	1.13	0.02	0.84	1.05	1.34	3.46	0.40
Cost of invested capital	3.6%	0.29	0.00	0.18	0.27	0.35	1.37	0.55
Total economic cost	**100.0%**	**7.96**	**2.48**	**6.14**	**7.48**	**8.98**	**24.55**	**0.36**

Table 4 - Outcome of the predicting methods with and without feature selection

		Multivariate Linear Regression (MLR)		K-Nearest-Neighbors (KNN)		Support Vector Regression (SVR)		Multilayer Perceptron (MLP)		Random Forest (RF)		Gradient Boosting (GB)		XGBoost	
	Extraction	*R2 score*	# features	*R2 score*	# features	*R2 score*	# features	*R2 score*	# features	*R2 score*	# features	*R2 score*	# features	*R2 score*	# features
***Mean* Model [*without* Feature Selection]**	**0**	0.685	39	0.514	39	0.728	39	0.737	39	0.792	39	0.722	39	0.808	39
	1	0.749	39	0.702	39	0.885	39	0.911	39	0.896	39	0.934	39	0.927	39
	2	0.804	39	0.471	39	0.804	39	0.884	39	0.771	39	0.851	39	0.872	39
	3	0.741	39	0.441	39	0.791	39	0.902	39	0.657	39	0.836	39	0.763	39
	4	0.759	39	0.702	39	0.909	39	0.906	39	0.792	39	0.826	39	0.902	39
	mean	**0.748**	**39**	**0.566**	**39**	**0.823**	**39**	**0.868**	**39**	**0.782**	**39**	**0.834**	**39**	**0.854**	**39**
	std.dev	**0.043**	**0**	**0.127**	**0**	**0.073**	**0**	**0.074**	**0**	**0.085**	**0**	**0.076**	**0**	**0.068**	**0**
***Mean* Model [*with* Feature Selection] Best Test Set**	**0**	0.755	7	0.817	8	0.846	10	0.762	5	0.832	5	0.834	6	0.818	11
	1	0.91	9	0.869	6	0.924	7	0.923	9	0.929	8	0.936	11	0.922	10
	2	0.822	8	0.743	5	0.88	5	0.927	7	0.771	4	0.902	5	0.860	7
	3	0.793	10	0.831	6	0.908	7	0.918	9	0.73	7	0.891	11	0.848	10
	4	0.876	11	0.850	10	0.924	10	0.900	11	0.829	11	0.876	7	0.903	7
	mean	**0.831**	**9**	**0.822**	**7**	**0.896**	**7.8**	**0.886**	**8.2**	**0.818**	**7**	**0.888**	**8**	**0.870**	**9**
	std.dev	**0.062**	**1.58**	**0.048**	**2**	**0.033**	**2.17**	**0.070**	**2.28**	**0.075**	**2.74**	**0.037**	**2.83**	**0.042**	**1.87**

Legend:

🟥	$R^2 < 0.60$
🟨	$0.60 \leq R^2 < 0.70$
🟧	$0.70 \leq R^2 < 0.75$
⬜	$0.75 \leq R^2 < 0.80$
🟩	$0.80 \leq R^2 < 0.85$
🟩	$0.85 \leq R^2 < 0.90$
🟩	$R^2 \geq 0.90$

Table 5 - Proposed cost models and related selected features

	SVR-e0-f10	GB-e1-f11	MLP-e2-f7	MLP-e3-f9	SVR-e4-f10	MLP-e3-f11
Average speed	True	True	/	True	True	True
Vehicle-revenue miles	/	True	/	/	/	/
Unlinked passenger trips	/	True	/	/	/	/
Passenger miles	/	True	/	/	/	/
Directional route miles	/	/	/	/	/	True
Average passengers per vehicle	True	/	/	True	True	True
Load factor	/	/	/	True	/	True
Network turnover	/	/	/	/	/	True
Average salary expenses per employee	True	True	True	True	True	True
Average hourly wage of transport operators	True	/	/	/	True	/
Operators productivity (annual miles)	True	True	True	True	True	True
Operators productivity (annual seats miles)	True	True	True	/	True	/
Percentage of transport operators (i.e., drivers and movement personnel)	True	True	True	True	True	True
Number of active fleet vehicles	/	True	/	/	/	/
Average fleet length	/	/	True	/	/	/
Vehicles productivity (annual miles)	True	/	True	True	True	True
Average lifetime miles of active fleet	/	/	/	True	/	True
Percentage of hybrid diesel vehicles	/	True	/	/	/	/
Percentage of electric vehicles	/	True	True	/	/	/
Fleet ownership cost (depreciation and cost of invested capital) per vehicle	True	/	/	True	True	True
Fleet ownership cost (depreciation and cost of invested capital) per seat (including standing)	True	/	/	/	True	/

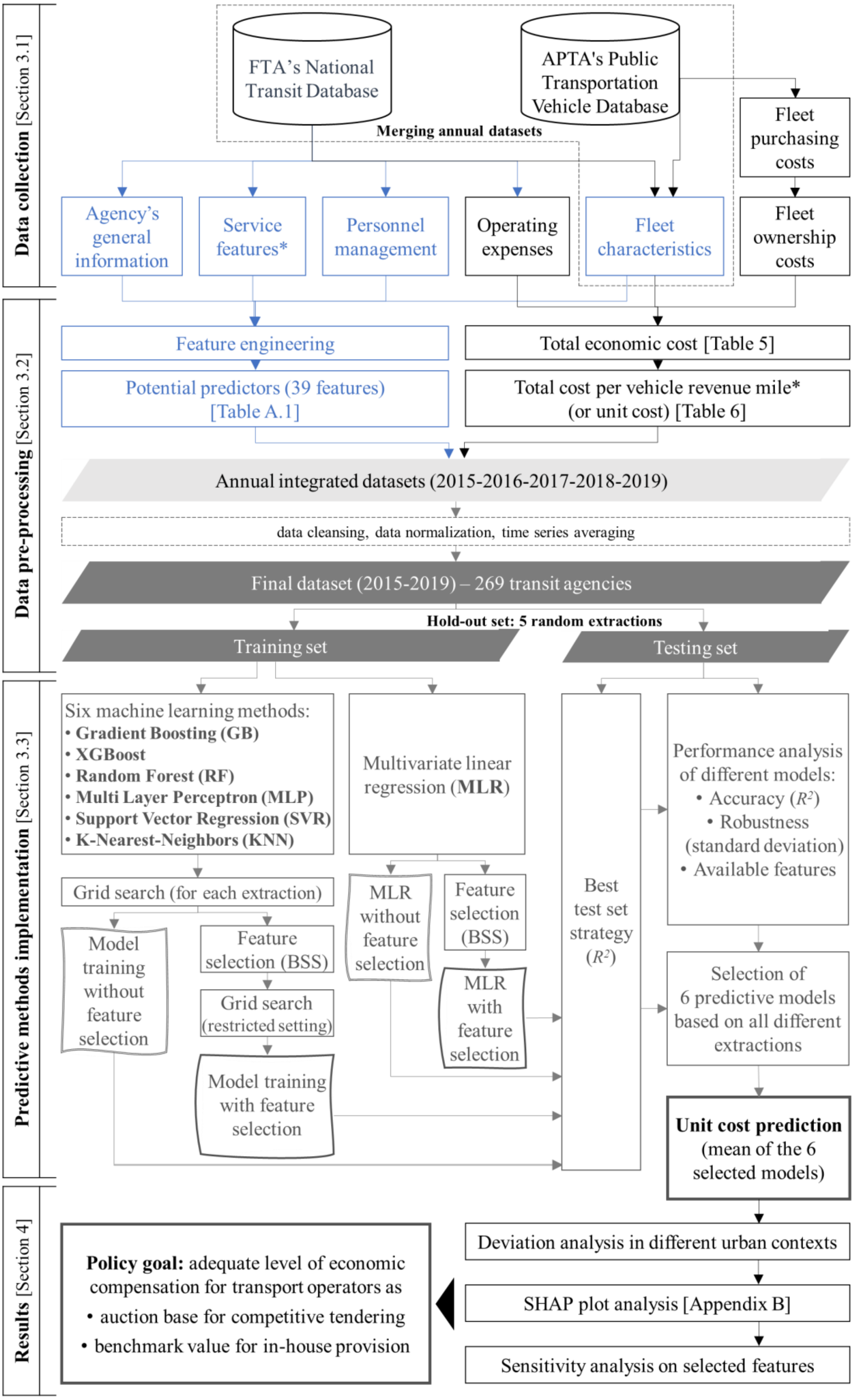
Table 6 - Features of LPBT services and unit costs predicted by the proposed models

City	Lancaster	Sheboygan	Tucson	Washington	Cleveland	Scranton
Primary UZA Population	402,004	71,313	843,168	4,586,770	1,780,673	381,502
Average speed	12.78	14.37	11.96	9.80	11.34	11.22
Vehicle-revenue miles	3,138,121	599,904	8,458,300	36,511,319	12,157,936	903,318
Deadhead miles	145,787	6,205	1,202,072	10,790,660	1,830,857	94,247
Directional route miles	736.00	85.00	1,083.00	2,132.18	1,231.50	291.00
Average passengers per vehicle	6.38	2.38	8.77	9.82	7.56	5.05
Network turnover	4,263.75	7,057.69	7,810.06	17,123.94	9,872.46	3,104.19
Average salary expenses per employee	45,211.04	35,060.77	43,922.60	70,985.51	54,855.27	41,557.80
Average hourly wage of transport operators	24.72	22.19	21.63	36.69	25.47	22.70
Operators productivity (annual miles)	19,123.22	17,961.20	18,111.99	11,434.80	14,673.04	14,569.65
Percentage of transport operators (i.e., drivers and movement personnel)	79.2%	77.5%	72.6%	69.0%	61.6%	68.9%
Fuel price (per gallon)	2.37	2.42	2.06	2.01	1.45	2.41
Number of active fleet vehicles	94	21	246	1,604	338	33
Average fleet age	6.91	11.10	9.15	8.78	7.85	5.45
Average fleet length	36.28	32.14	40.00	41.27	42.62	35.00
Average fleet capacity (seats + standing)	48.31	42.38	59.19	67.16	64.82	69.55
Vehicles productivity (annual miles)	33,172.89	29,547.38	39,751.66	28,783.44	41,070.14	32,546.61
Percentage of diesel vehicles	34.0%	100.0%	81.7%	12.1%	68.9%	21.2%
Percentage of CNG vehicles	0.0%	0.0%	18.3%	33.9%	31.1%	39.4%
Percentage of hybrid diesel vehicles	66.0%	0.0%	0.0%	53.9%	0.0%	39.4%
Percentage of electric vehicles	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%
Fleet ownership cost (depreciation and cost of invested capital) per vehicle	55,396.67	27,484.80	37,805.13	49,527.55	44,182.01	54,570.84
Total cost per vrm [USD/vrm]* (observed)	**6.75**	**4.71**	**6.94**	**17.45**	**11.56**	**8.83**
Total cost per vrm predicted by the proposed models [USD/vrm]						
SVR-e0-f10	7.08	5.34	6.83	14.89	10.26	8.80
GB-e1-f11	6.97	5.66	7.39	16.96	11.07	8.13
MLP-e2-f7	6.53	5.80	7.04	17.60	10.65	7.93
MLP-e3-f9	6.95	5.39	6.90	16.75	10.49	8.80
SVR-e4-f10	7.13	5.25	6.89	14.93	10.36	8.79
MLP-e3-f11	7.00	5.34	6.89	16.21	10.56	8.97
Mean	**6.94**	**5.46**	**6.99**	**16.22**	**10.56**	**8.57**
% error w.r.t. observed unit cost	2.8%	16.0%	0.7%	-7.0%	-8.6%	-3.0%
standard deviation	0.21	0.21	0.21	1.11	0.28	0.43

* Since the LPBT services are produced in a period ranging from 2015 up to 2019, the observed monetary values are given in 2019 prices (in USD), consistently with the unit costs predicted by the proposed models.

Figures

Figure 1 – The analytical framework developed to predict the unit cost of LPBT services



Data collection [Section 3.1]
Data pre-processing [Section 3.2]
Predictive methods implementation [Section 3.3]
Results [Section 4]
FTA's National Transit Database
APTA's Public Transportation Vehicle Database
Merging annual datasets
Fleet purchasing costs
Agency's general information
Service features*
Personnel management
Operating expenses
Fleet characteristics
Fleet ownership costs
Feature engineering
Potential predictors (39 features) [Table A.1]
Total economic cost [Table 5]
Total cost per vehicle revenue mile* (or unit cost) [Table 6]
Annual integrated datasets (2015-2016-2017-2018-2019)
data cleansing, data normalization, time series averaging
Final dataset (2015-2019) – 269 transit agencies
Hold-out set: 5 random extractions
Training set
Testing set
Six machine learning methods:
• Gradient Boosting (GB)
• XGBoost
• Random Forest (RF)
• Multi Layer Perceptron (MLP)
• Support Vector Regression (SVR)
• K-Nearest-Neighbors (KNN)
Multivariate linear regression (MLR)
Grid search (for each extraction)
Model training without feature selection
Feature selection (BSS)
Grid search (restricted setting)
Model training with feature selection
MLR without feature selection
Feature selection (BSS)
MLR with feature selection
Best test set strategy (R^2)
Performance analysis of different models:
• Accuracy (R^2)
• Robustness (standard deviation)
• Available features
Selection of 6 predictive models based on all different extractions
Unit cost prediction (mean of the 6 selected models)
Policy goal: adequate level of economic compensation for transport operators as
• auction base for competitive tendering
• benchmark value for in-house provision
Deviation analysis in different urban contexts
SHAP plot analysis [Appendix B]
Sensitivity analysis on selected features

Figure 2 – Total cost per vrm as "core" service features increase/decrease

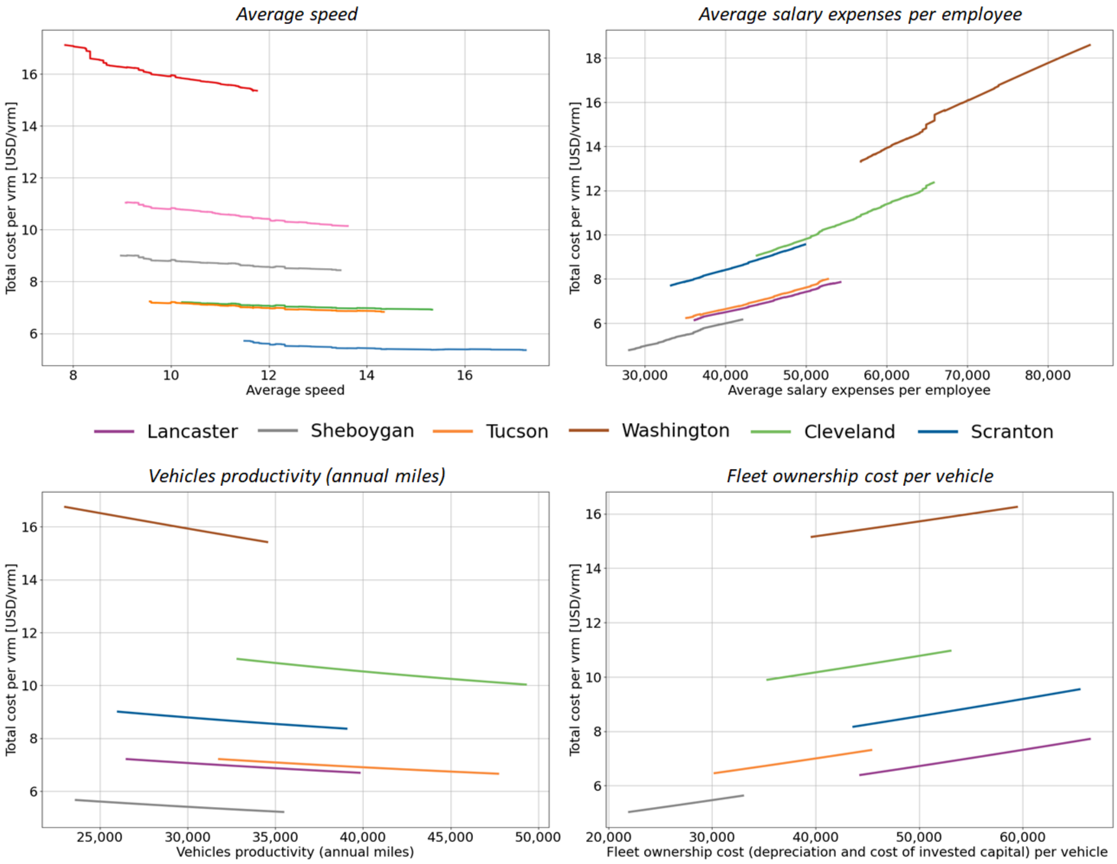
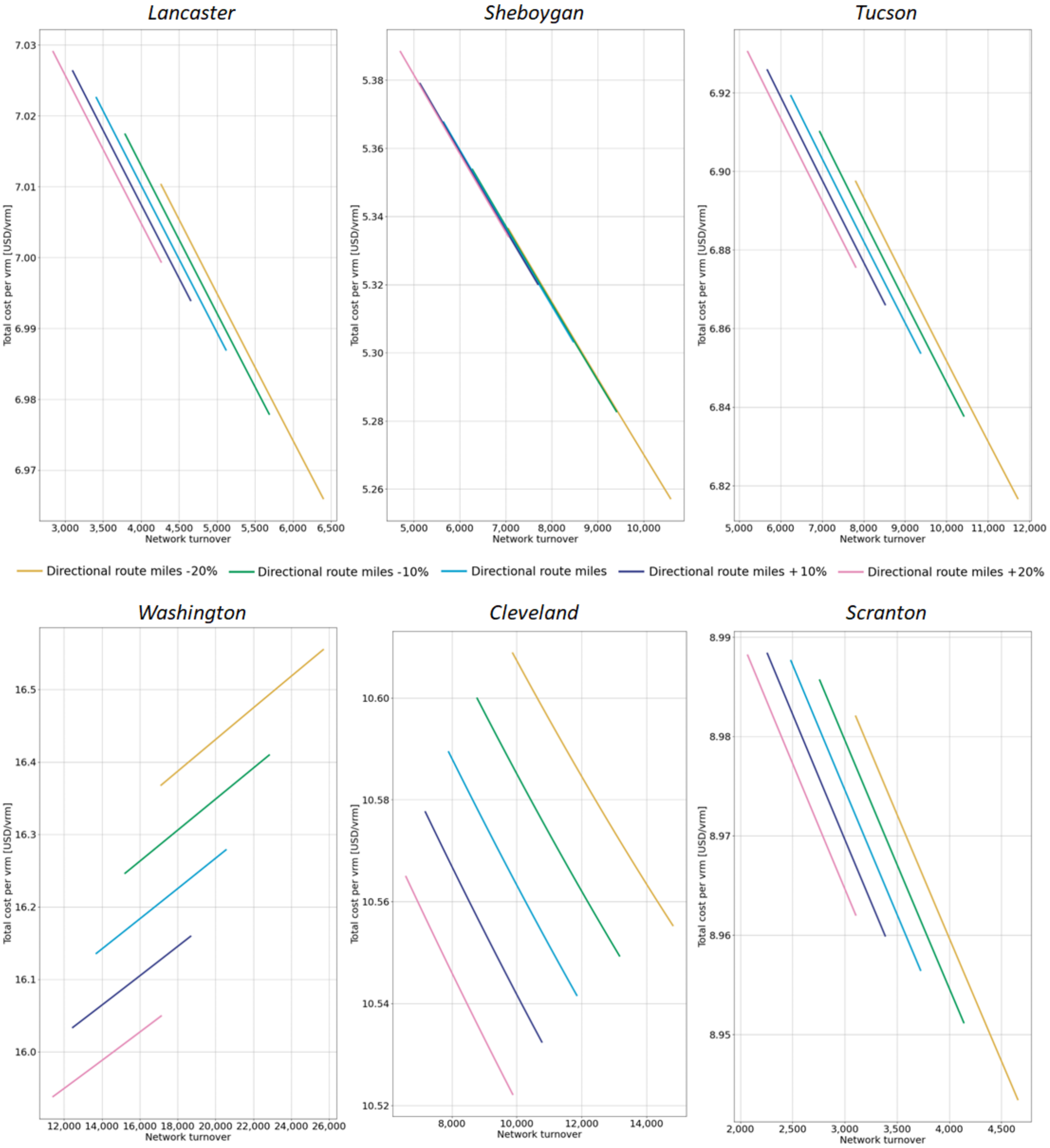
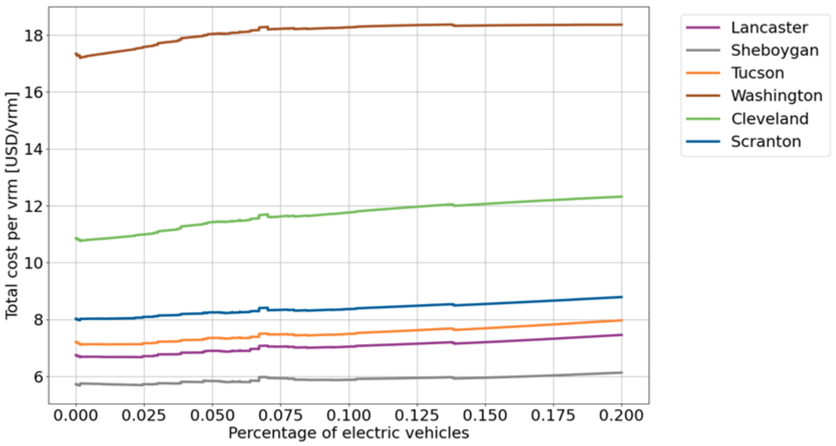


Figure 3 – Total cost per vrm as two important policy features increase/decrease

- **Network turnover (and directional route miles)**



- **Percentage of electric vehicles**



Appendix A

In this document additional tables are reported for the methodology section (Section 4).

Table A.1 - Some descriptive statistics related to service features of transit agencies included in the sample

	Mean	*Min*	*1° quartile*	*Median*	*3° quartile*	*Max*	*Coeff. of variation*
Vehicle-revenue miles (mi)	5,222,623	29,272	848,222	1,724,047	4,238,707	87,657,339	2.00
Directional route miles (mi)	533	3.60	160	307	599	5,616	1.22
Unlinked Passenger Trips	14,434,387	5,741	1,032,444	2,748,108	7,978,861	743,763,755	3.68
Fuel Used (gallon equivalent)	1,644,671	5,286	197,255	423,389	1,303,923	37,704,550	2.42
Average speed (mi/h)	13.29	4.71	11.79	13.05	14.56	27.77	0.20

Table A.2 - Some descriptive statistics related to personnel management of transit agencies included in the sample

	Mean	*Min*	*1° quartile*	*Median*	*3° quartile*	*Max*	*Coeff. of variation*
Number of total employees	529	3	69	150	383	14,918	2.48
Percentage of full time employees	0.84	0.00	0.78	0.92	0.98	1.00	0.24
Average salary expenses per employee (USD)	44,035.64	6,962.35	36,118.38	44,661.82	51,748.16	82,002.74	0.29
Percentage of transport operators (i.e., drivers and movement personnel)	361	2	52	109	269	9,883	2.40
Number of hours worked by transport operators	680,314	4,171	89,653	194,168	519,348	20,362,247	2.49
Average hourly wage of transport operators (USD)	23.69	9.21	19.91	23.52	27.02	53.33	0.25

Table A.3 - Some descriptive statistics related to fleet characteristics of transit agencies included in the sample

	Mean	*Min*	*1° quartile*	*Median*	*3° quartile*	*Max*	*Coeff. of variation*
Number of active fleet vehicles	174	1	30	58	142	3,964	2.19
Average fleet age	8.70	1.00	7.28	8.55	10.04	22.50	0.27
Average fleet length (feet)	36.48	17.00	33.65	36.63	39.63	52.23	0.12
Average fleet capacity (seats + standing)	56	5	47	56	64	118	0.26
Vehicles productivity (annual miles)	33,095	6,178	26,762	32,729	39,065	72,145	0.28
Average lifetime miles of active fleet	283,976	11,730	212,457	276,850	335,549	3,295,098	0.51
Percentage of diesel vehicles	0.68	0.00	0.48	0.80	0.99	1.00	0.48
Percentage of CNG vehicles	0.15	0.00	0.00	0.00	0.14	1.00	1.92
Percentage of hybrid diesel vehicles	0.09	0.00	0.00	0.00	0.11	1.00	1.82
Percentage of electric vehicles	0.01	0.00	0.00	0.00	0.00	1.00	4.19

Table A.4 - Features included in the analysis

Major category	***Features included in the analysis***	
General information	1	Population of primary urbanized area served
Service features	2	Average speed
	3	Vehicle miles
	4	Vehicle-revenue miles
	5	Dedhead miles $[= 3 - 4]$
	6	Seat revenue miles (including standing) $[= 4 \times 28]$
	7	Directional route miles
	8	Network turnover $[= 4/7]$
	9	Unlinked passenger trips
	10	Passenger miles
	11	Average passengers per vehicle $[= 10/4]$
	12	Load factor $[= 11/28]$
	13	Fuel used (gallon equivalent)
	14	Fuel price (per gallon) $[= (OPEX\ related\ to\ fuel/energy)/13]$
Personnel management	15	Number of total employees
	16	Percentage of full-time employees
	17	Average salary expenses per employee
	18	Number of transport operators (i.e., drivers and movement personnel)
	19	Percentage of transport operators (i.e., drivers and movement personnel) $[= 18/15]$
	20	Number of hours worked by transport operators
	21	Average hourly wage of transport operators
	22	Operators productivity (annual operations hours) $[= 20/18]$
	23	Operators productivity (annual miles) $[= 4/18]$
	24	Operators productivity (annual seats miles) $[= 6/18]$
Fleet characteristics	25	Number of active fleet vehicles
	26	Average fleet age
	27	Average fleet length
	28	Average fleet capacity (seats + standing)
	29	Vehicles productivity (annual miles)
	30	Average lifetime miles of active fleet
	31	Percentage of diesel vehicles
	32	Percentage of CNG vehicles
	33	Percentage of hybrid diesel vehicles
	34	Percentage of electric vehicles
	35	Percentage of hydrogen vehicles
	36	Percentage of other powertrains
	37	Percentage of low carbon buses $[= 33 + 34 + 35]$
Economic costs	38	Fleet ownership cost (depreciation and cost of invested capital) per vehicle $[= (Vehicle\ depreciation\ costs + Cost\ of\ invested\ capita)/25]$
	39	Fleet ownership cost (depreciation and cost of invested capital) per seat (including standing) $[= (Vehicle\ depreciation\ costs + Cost\ of\ invested\ capita)/(25 \times 28)]$

Table A.5 - Hyperparameter space used to tune the ML methods with Grid Search*

Gradient Boosting (GB)	
hyperparam. name	grid values
learning_rate	1, 0.1, 0.01, 0.05
max_depth	2,3,4,5,6
min_samples_split	2,5,10,15,20
subsample	0.5,0.75,1
n_estimators	100, 500, 1000, 2000

Multilayer Perceptron (MLP)	
hyperparam. name	grid values
activation	relu, logistic, tanh
alpha	0.1, 0.5, 0.05
hidden_layer_sizes	(25,), (50,), (100,), (25,25), (50,25), (50,50)
max_iter	100,200,500,1000
solver	adam, lbfgs

Support Vector Regression (SVR)	
hyperparam. name	grid values
C	1, 10, 100, 1000
gamma	0.1, 0.01, 0.001, 0.0001

XGBoost	
hyperparam. name	grid values
learning_rate	0.1, 0.05
max_depth	2, 5
n_estimators	100, 500, 1000

Random Forest (RF)	
hyperparam. name	grid values
max_depth	2,5,10,20
min_samples_split	2,5,10
n_estimators	100,500,1000

K-Nearest-Neighbors (KNN)	
hyperparam. name	grid values
weights	uniform, distance
n_neighbors	from 1 to 15
algorithm	auto, ball_tree, kd_tree, brute
leaf_size	1,2,3,4,5,10,15,20,25,30
p	1,2,3

* For each of the used ML method, we have presented the tuned parameters, the other not listed in the table are the default ones (except for GB loss=huber and for MLP learning_rate=adaptive, early_stopping=True).

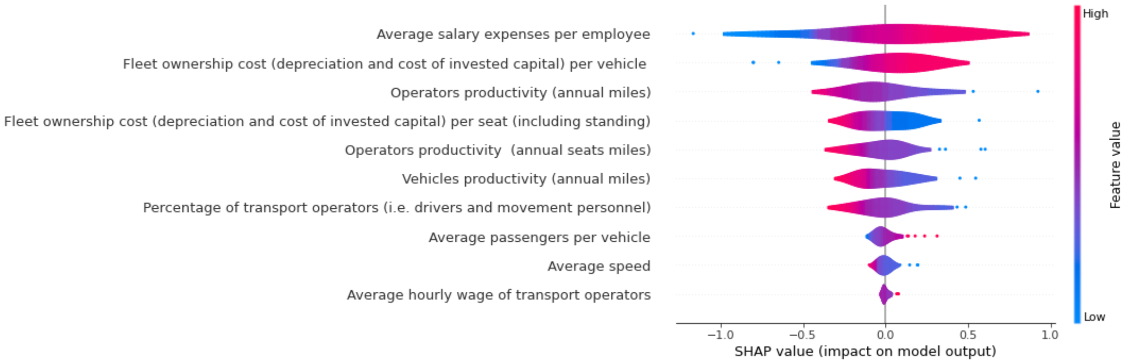
Table A.1 - The parameters used in the ML models selected

ML model	**Number of features**	**Extraction**
Support Vector Regression (SVR)	10	0
Parameters C: *100* Gamma: *0.001*		
Gradient Boosting (GB)	11	1
Parameters learning_rate: *0.05* max_depth: *2* min_samples_split: *2* Subsample: *0.5* n_estimators: *1000*		
Multilayer Perceptron (MLP)	7	2
Parameters activation: *logistic* alpha: *0.05* hidden_layer_sizes: *(50,25)* max_iter: *100* solver: *lbfgs*		
Multilayer Perceptron (MLP)	9	3
Parameters activation: *logistic* alpha: *0.5* hidden_layer_sizes: *(100,)* max_iter: *500* solver: *lbfgs*		
Support Vector Regression (SVR)	10	4
Parameters C: *100* Gamma: *0.001*		
Multilayer Perceptron (MLP)	11	3
Parameters activation: *relu* alpha: *0.5* hidden_layer_sizes: *(100,)* max_iter: *500* solver: *lbfgs*		

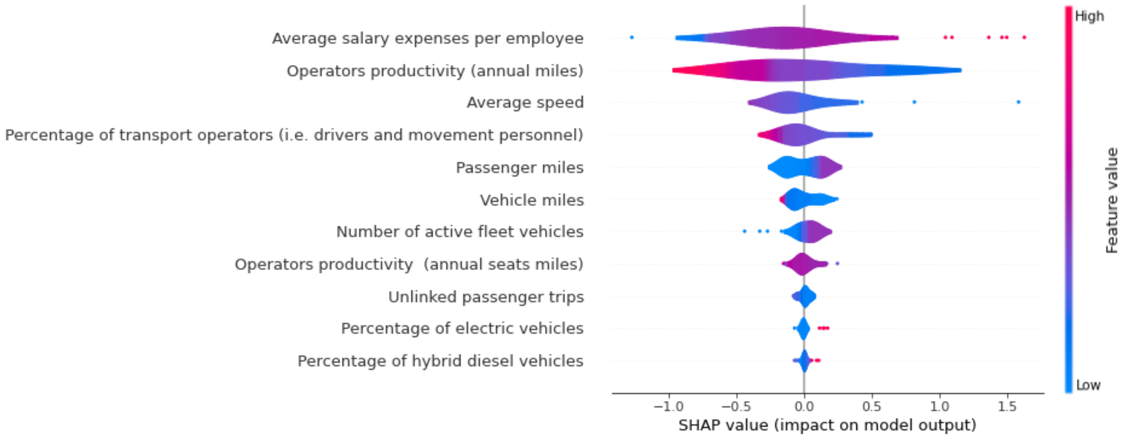
Appendix B

This document displays the SHAP plots for each proposed model. A local regression using the Kernel SHAP function was performed for each model in the best test set. The final output, presented as violin plots, provides information on the contribution of individual features in terms of the direction and scale of their impact on the model outcome. In each summary plot, features are listed on the y-axis in order of importance from top to bottom, along with their mean SHAP values. The x-axis displays the SHAP values, indicating the degree of influence each feature has on the model output (positive or negative). A wider SHAP value (wider violin plot) signifies a greater impact on the model outcome. The color gradient illustrates the direction of this impact, with red indicating high values and blue indicating low values.

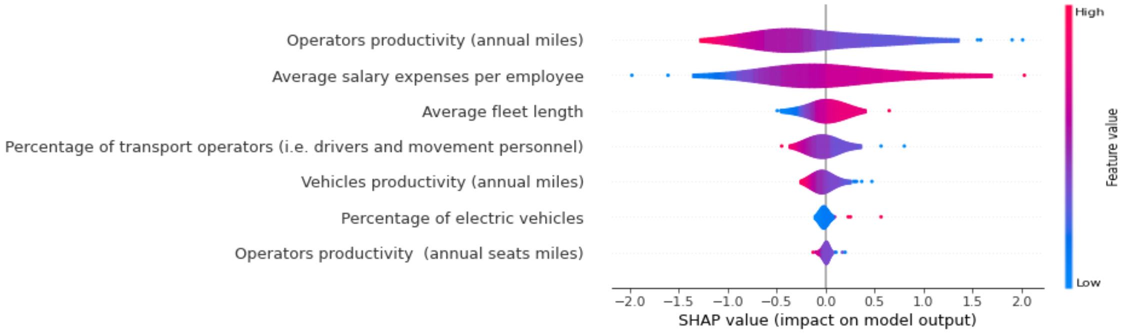
Model: SVR-e0-f10



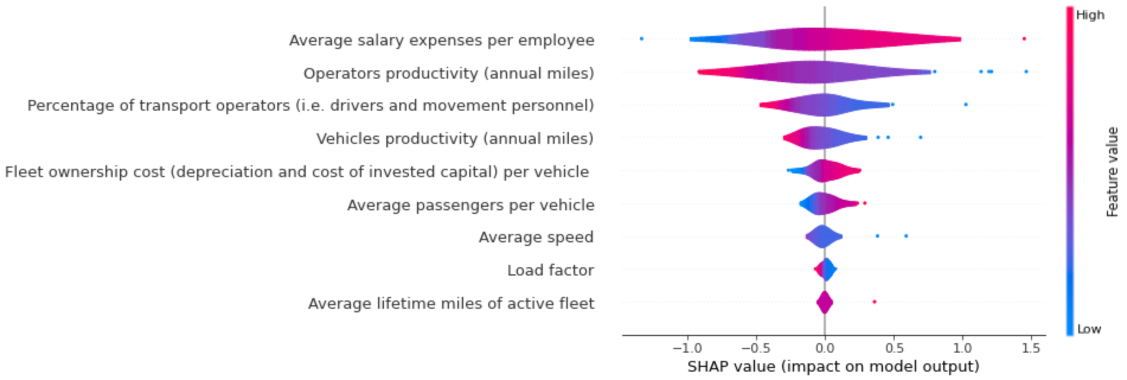
Model: GB-e1-f11



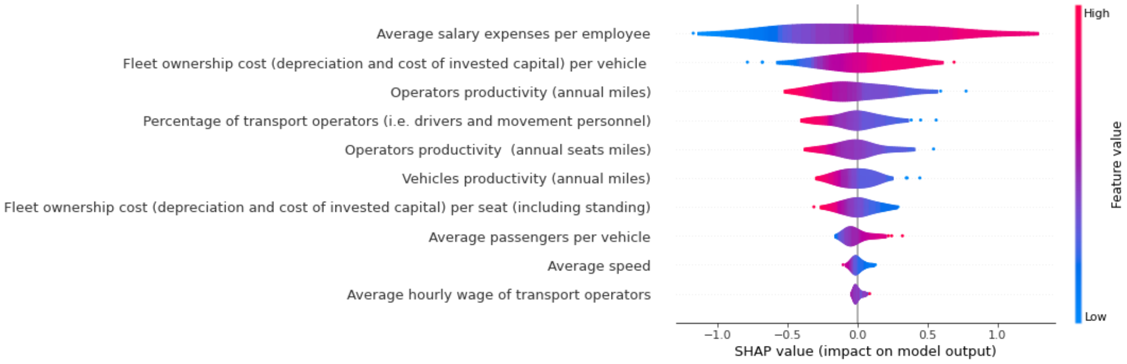
Model: MLP-e2-f7



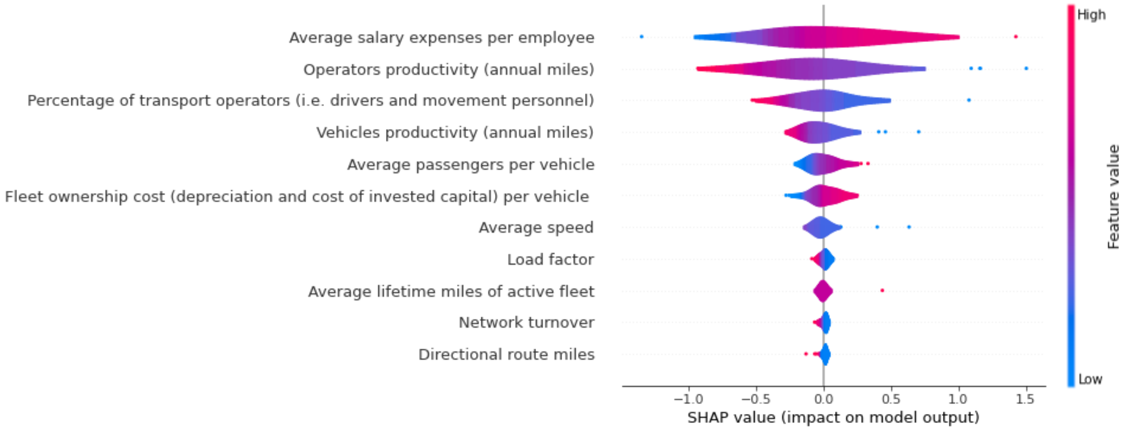
Model: MLP-e3-f9



Model: SVR-e4-f10



Model: MLP-e3-f11



References

Amaral M, Saussier S, Yvrande-Billon A. Auction procedures and competition in public services: The case of urban public transport in France and London. *Utilities Policy* 2009, 17(2): 166-175.

Avenali, A., Boitani, A., Catalano, G., D’Alfonso, T., & Matteucci, G. (2016). Assessing standard costs in local public bus transport: Evidence from Italy. *Transport Policy*, 52, 164-174.

Avenali, A., Boitani, A., Catalano, G., D’Alfonso, T., & Matteucci, G. (2018). Assessing standard costs in local public bus transport: A hybrid cost model. *Transport Policy*, 62, 48-57.

Ban, T., Zhang, R., Pang, S., Sarrafzadeh, A., & Inoue, D. (2013, November). Referential knn regression for financial time series forecasting. In *International Conference on Neural Information Processing* (pp. 601-608). Springer, Berlin, Heidelberg.

Berechman, J. (1987). Cost structure and production technology in transit: An application to the Israeli bus transit sector. *Regional Science and Urban Economics*, 17(4), 519-534.

Bertsekas, D. P., Hager, W., & Mangasarian, O. (1999). Nonlinear programming. athena scientific belmont. Massachusets, U.S..

Bhattacharyya, A., Kumbhakar, S. C., & Bhattacharyya, A. (1995). Ownership structure and cost efficiency: A study of publicly owned passenger-bus transportation companies in *India. Journal of Productivity Analysis*, 6(1), 47-61.

Bhavsar, P., Safro, I., Bouaynaya, N., Polikar, R., & Dera, D. (2017). Machine learning in transportation data analytics. In *Data analytics for intelligent transportation systems (pp. 283-307)*. Elsevier.

Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford university press.

Bishop, C. M., & Nasrabadi, N. M. (2006). Pattern recognition and machine learning. New York: springer.

Bodendorf, F., Merkl, P., & Franke, J. (2021). Intelligent cost estimation by machine learning in supply management: A structured literature review. *Computers & Industrial Engineering*, 160, 107601.

Boitani, A., Nicolini, M., & Scarpa, C. (2013). Do competition and ownership matter? Evidence from local public transport in Europe. *Applied economics*, 45(11), 1419-1434.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152).

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). Classification and regression trees. Routledge.

Cambini, C., Filippini, M., 2003. Competitive tendering and optimal size in the regional bus transportation industry: An example from Italy. *Annals of Public and Cooperative Economics*, 74(1): 163–182

Cambini, C., Piacenza, M., & Vannoni, D. (2007). Restructuring public transit systems: evidence on cost properties from medium and large-sized companies. *Review of Industrial Organization*, 31(3), 183-203.

Catal, C. (2011). Software fault prediction: A literature review and current trends. *Expert systems with applications*, 38(4), 4626-4636.

Ceccarelli, G., Cantelmo, G., Nigro, M., & Antoniou, C. (2021, June). Machine Learning from imbalanced datasets: an application to the bike-sharing inventory problem. In *2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems* (MT-ITS) (pp. 1-6). IEEE.

Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology* (TIST), 2(3), 1-27.

Chelazzi, C., Villa, G., Manno, A., Ranfagni, V., Gemmi, E., & Romagnoli, S. (2021). The new SUMPOT to predict postoperative complications using an Artificial Neural Network. *Scientific reports*, 11(1), 1-12.

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* (pp. 785-794).

Chen, Y., Zhang, Y., & Sun, R. (2021). Data-driven estimation of energy consumption for electric bus under real-world driving conditions. *Transportation Research Part D: Transport and Environment*, 98, 102969.

Chou, J. S. (2009). Generalized linear model-based expert system for estimating the cost of transportation projects. *Expert Systems with Applications*, 36(3), 4253-4267.

Colburn, C. B. & W. K. Talley (1992): A firm specific analysis of economies of size in the U.S. urban multiservice transit industry, *Transportation Research*, Part B(3), 195–206.

Comello, S., Glenk, G., & Reichelstein, S. (2021). Transitioning to clean energy transportation services: Life-cycle cost analysis for vehicle fleets. *Applied Energy*, 285, 116408.

Dalen, D. M., & Gomez-Lobo, G. L. (1996). Regulation and incentive contracts: An empirical investigation of the Norwegian bus transport industry (No. W96/08). Institute for Fiscal Studies.

Dalen, D. M., & Gómez-Lobo, A. (2003). Yardsticks on the road: Regulatory contracts and cost efficiency in the Norwegian bus industry. *Transportation*, 30, 371-386.

Damodaran, A. (2012). Investment valuation: Tools and techniques for determining the value of any asset (Vol. 666). John Wiley & Sons.

Damodaran A. (2021). Discount rate estimation – Cost of Capital by Industry Sector, http://people.stern.nyu.edu/adamodar/New_Home_Page/datafile/wacc.html

Daraio, C., Diana, M., Di Costa, F., Leporelli, C., Matteucci, G., & Nastasi, A. (2016). Efficiency and effectiveness in the urban public transport sector: A critical review with directions for future research. *European Journal of Operational Research*, 248(1), 1-20.

Dietterich, T. G. (2000, June). Ensemble methods in machine learning. In International workshop on multiple classifier systems (pp. 1-15). Springer, Berlin, Heidelberg.

Fan, R. E., Chang, K. W., Hsieh, C. J., Wang, X. R., & Lin, C. J. (2008). LIBLINEAR: A library for large linear classification. the *Journal of machine Learning research*, 9, 1871-1874.

Farsi, M., Fetz, A., & Filippini, M. (2007). Economies of scale and scope in local public transportation. *Journal of Transport Economics and Policy* (JTEP), 41(3), 345-361.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.

Filippini, M., & Prioni, P. (1994). Is scale and cost inefficiency in the Swiss bus industry a regulatory problem? Evidence from a frontier cost approach. *Journal of the Economics of Business*, 1(2), 219-232.

Filippini, M., & Prioni, P. (2003). The influence of ownership on the cost of bus service provision in Switzerland-an empirical illustration. *Applied Economics*, 35(6), 683-690.

Fraquelli, G., Piacenza, M., & Abrate, G. (2001). Il trasporto pubblico locale in Italia: variabili esplicative dei divari di costo tra le imprese. *Economia e Politica industriale*.

Fraquelli, G., Piacenza, M., & Abrate, G. (2004). Regulating Public Transit Networks: How do Urban-Intercity Diversification and Speed-up Measures Affect Firms' Cost Performance?. *Annals of public and Cooperative Economics*, 75(2), 193-225.

Gagnepain, P., & Ivaldi, M. (2002). Incentive regulatory policies: the case of public transit systems in France. *RAND Journal of Economics*, 605-629.

Gammelli, D., Peled, I., Rodrigues, F., Pacino, D., Kurtaran, H. A., & Pereira, F. C. (2020). Estimating latent demand of shared mobility through censored gaussian processes. *Transportation Research Part C: Emerging Technologies*, 120, 102775.

Grippo, L., Manno, A., & Sciandrone, M. (2015). Decomposition techniques for multilayer perceptron training. *IEEE transactions on neural networks and learning systems*, 27(11), 2146-2159.

Gunduz, M., Ugur, L. O., & Ozturk, E. (2011). Parametric cost estimation system for light rail transit and metro trackworks. *Expert Systems with Applications*, 38(3), 2873-2877.

Hagenauer, J., & Helbich, M. (2017). A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Systems with Applications*, 78, 273-282.

Hashemi, S. T., Ebadati, O. M., & Kaur, H. (2020). Cost estimation and prediction in construction projects: a systematic review on machine learning techniques. *SN Applied Sciences*, 2(10), 1-27.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

Hensher, D. A., & Button, K. J. (2000). Handbook of transport modelling (No. 1).

Hensher DA, Wallis IP. Competitive tendering as a contracting mechanism for subsidising transport: the bus experience. *Journal of Transport Economics and Policy* 2005, 39 (3): 295-322.

Ho, T. K. (1995, August). Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278-282). IEEE.

Huang, J., Li, Y. F., & Xie, M. (2015). An empirical analysis of data preprocessing for machine learning-based software cost estimation. *Information and software Technology*, 67, 108-127.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.

Kohli, S., Godwin, G. T., & Urolagin, S. (2021). Sales prediction using linear and KNN regression. In *Advances in Machine Learning and Computational Intelligence* (pp. 321-329). Springer, Singapore.

Karlaftis, M., & McCarthy, P. (1999). The effect of privatization on public transit costs. *Journal of Regulatory Economics*, 16(1), 27-44.

Kurasova, O., Marcinkevičius, V., Medvedev, V., & Mikulskienė, B. (2021). Early cost estimation in customized furniture manufacturing using machine learning. *International Journal of Machine Learning and Computing*, 11(1), 28-33.

Laboissiere, L. A., Fernandes, R. A., & Lage, G. G. (2015). Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks. *Applied Soft Computing*, 35, 66-74.

Laver, R., Schneck, D., Skorupski, D., Brady, S., & Cham, L. (2007). Useful life of transit buses and vans (No. FTA-VA-26-7229-07.1).

Leshno, M., Lin, V. Y., Pinkus, A., & Schocken, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural networks*, 6(6), 861-867.

Li, H., Parikh, D., He, Q., Qian, B., Li, Z., Fang, D., & Hampapur, A. (2014). Improving rail network velocity: A machine learning approach to predictive maintenance. *Transportation Research Part C: Emerging Technologies*, 45, 17-26.

Li, W., Yin, Y., Quan, X., & Zhang, H. (2019). Gene expression value prediction based on XGBoost algorithm. *Frontiers in genetics*, 1077.

Liu, Z., Liu, Y., Meng, Q., & Cheng, Q. (2019). A tailored machine learning approach for urban transport network flow estimation. *Transportation Research Part C: Emerging Technologies*, 108, 130-150.

Liu, Y., Zhang, Q., Lyu, C., & Liu, Z. (2021). Modelling the energy consumption of electric vehicles under uncertain and small data conditions. *Transportation Research Part A: Policy and Practice*, 154, 313-328.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

Ma, T., Antoniou, C., & Toledo, T. (2020). Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast. *Transportation Research Part C: Emerging Technologies*, 111, 352-372.

Mahpour, A., & El-Diraby, T. (2022). Application of Machine-Learning in Network-Level Road Maintenance Policy-Making: The Case of Iran. *Expert Systems with Applications*, 191, 116283.

Manno, A., Martelli, E., & Amaldi, E. (2022). A Shallow Neural Network Approach for the Short-Term Forcast of Hourly Energy Consumption. *Energies*, 15(3), 958.

Manno, A., Palagi, L., & Sagratella, S. (2018). Parallel decomposition methods for linearly constrained problems subject to simple bound with application to the SVMs training. *Computational Optimization and Applications*, 71(1), 115-145.

Manno, A., Rossi, F., Smriglio, S., & Cerone, L. (2021). Comparing Deep and Shallow Neural Networks in Forecasting Call Center Arrivals. https://www.researchsquare.com/article/rs-670306/v1.pdf, 2021.

Manno, A., Sagratella, S., & Livi, L. (2016, July). A convergent and fully distributable SVMs training algorithm. In *2016 International Joint Conference on Neural Networks* (IJCNN) (pp. 3076-3080). IEEE.

Marković, N., Kim, M. E., & Schonfeld, P. (2016). Statistical and machine learning approach for planning dial-a-ride systems. *Transportation Research Part A: Policy and Practice*, 89, 41-55.

Pan, B. (2018, February). Application of XGBoost algorithm in hourly PM2. 5 concentration prediction. In IOP conference series: earth and environmental science (Vol. 113, No. 1, p. 012127). IOP publishing.

Piacenza, M. (2006). Regulatory contracts and cost efficiency: Stochastic frontier evidence from the Italian local public transport. *Journal of Productivity Analysis*, 25(3), 257-277.

Plakandaras, V., Papadimitriou, T., & Gogas, P. (2019). Forecasting transportation demand for the US market. Transportation Research Part A: Policy and Practice, 126,

Qasim, M., & Khadkikar, V. (2014). Application of artificial neural networks for shunt active power filter control. *IEEE Transactions on industrial informatics*, 10(3), 1765-1774.

Raju, N., Arkatkar, S., Joshi, G., & Antoniou, C. (2022). Data-Driven Approach for Modeling the Mixed Traffic Conditions Using Supervised Machine Learning. In *Intelligent Infrastructure in Transportation and Management* (pp. 3-12). Springer, Singapore.195-214.

Ripplinger, D. G., & Bitzan, J. D. (2018). The cost structure of transit in small urban and rural US communities. *Transportation Research Part A: Policy and Practice*, 117, 176-189.

Roy, W., & Yvrande-Billon, A. (2007). Ownership, contractual practices and technical efficiency: The case of urban public transport in France. *Journal of Transport Economics and Policy* (JTEP), 41(2), 257-282.

Salas, P., De la Fuente, R., Astroza, S., & Carrasco, J. A. (2022). A systematic comparative evaluation of machine learning classifiers and discrete choice models for travel mode choice in the presence of response heterogeneity. *Expert Systems with Applications*, 193, 116253.

Servadei, L., Mosca, E., Zennaro, E., Devarajegowda, K., Werner, M., Ecker, W., & Wille, R. (2020). Accurate cost estimation of memory systems utilizing machine learning and solutions from computer vision for design automation. *IEEE Transactions on Computers*, 69(6), 856-867.

Schultz, M., & Reitmann, S. (2019). Machine learning approach to predict aircraft boarding. *Transportation Research Part C: Emerging Technologies*, 98, 391-408.

Shapley, L.S. (1951). Notes on the n-Person Game - II: The Value of an n-Person Game. Santa Monica, Calif.: RAND Corporation.

Shapley, L.S. (1953). “A value for n-person games”. In: Contributions to the Theory of Games, 2.28 (1953), Kuhn H.W. and Tucker A.W. editors, pp. 307–317, Princeton University Press, Princeton, New Jersey.

Shaw-Er, J., Chiang, W., & Chen, Y. W. (2005). Cost structure and technological change of local public transport: the Kaohsiung City Bus case. *Applied Economics*, 37(12), 1399-1410.

Zimek, A., Gaudet, M., Campello, R. J., & Sander, J. (2013, August). Subsampling for efficient and effective unsupervised outlier detection ensembles. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 428-436).

Shleifer, A. (1985). A theory of yardstick competition. The RAND journal of Economics, 319-327.

Tizghadam, A., Khazaei, H., Moghaddam, M. H., & Hassan, Y. (2019). Machine learning in transportation. *Journal of Advanced Transportation*, 2019.

Verdeaux, J. J. (2003). Public procurement in the European Union and in the United States: a comparative study. *Public Contract Law Journal*, 713-738.

Viton, P. A. (1981). A translog cost function for urban bus transit. *The Journal of Industrial Economics*, 287-304.

Yang, K., Zhao, W., & Antoniou, C. (2020, September). Utilizing Import Vector Machines to Identify Dangerous Pro-active Traffic Conditions. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems* (ITSC) (pp. 1-6). IEEE.

Yeh, T. H., & Deng, S. (2012). Application of machine learning methods to cost estimation of product life cycle. *International Journal of Computer Integrated Manufacturing*, 25(4-5), 340-352.

Viton, P. A. (1992): Consolidations of Scale and Scope in Urban Transit, *Regional Science and Urban Economics*, 22(1), 25–49.

Williams, M. (1979). Firm size and operating costs in urban bus transportation. *The journal of Industrial Economics*, 209-218.

Wunsch, P. (1996). Cost and productivity of major urban transit systems in Europe: an exploratory analysis. *Journal of Transport Economics and Policy*, 171-186.

Jacob, C., & Abdulhai, B. (2010). Machine learning for multi-jurisdictional optimal traffic corridor control. *Transportation Research Part A: Policy and Practice*, 44(2), 53-64.

Xu, C., Ji, J., & Liu, P. (2018). The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transportation research part C: emerging technologies*, 95, 47-60.

Zheng, H., & Wu, Y. (2019). A xgboost model with weather similarity analysis and feature engineering for short-term wind power forecasting. *Applied Sciences*, 9(15), 3019.