Elektronik Sosyal Bilimler Dergisi® www.esosder.org
Electronic Journal of Social Sciences® ISSN:1304-0278
Temmuz/July(2024) - Cilt/Volume:23 - Sayı/Issue:91 (1041-1057)

PREDICTION OF DRIVING TIME OF ELECTRIC SCOOTER (E-SCOOTER) DRIVERS BY MACHINE LEARNING

ELEKTRİKLİ SCOOTER (E-SCOOTER) SÜRÜCÜLERİNİN SÜRÜŞ SÜRESİNİN MAKİNE ÖĞRENMESİ İLE TAHMİNİ

Hakan İNAC[1]

Abstract

This study aims to estimate the driving times of drivers who prefer electric scooter vehicles. In general, e-scooters reduce the loss of time caused by traffic jams thanks to their smaller size and maneuverability, therefore, these vehicles provide rapid progress in urban journeys. E-scooters also offer an advantage in finding a parking space and easy parking thanks to their more compact structure. In this study, ML algorithms were used to predict the driving times of drivers who prefer e-scooter vehicles. The AB model performed well with a low Mean Square Error (MSE) value (0.005). The Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) values were also relatively low (0.069 and 0.039, respectively), indicating that the model's predictions were close to the actual values. Also, the high R-squared-Coefficient of Determination (R^2) value (0.947) suggested that this model explained the data quite well, and its predictions approached the actual values with high accuracy. On the other hand, the GB algorithm performed poorly compared to different algorithms, with its high margin of error and low accuracy rate. These results provide an advantage in time management by estimating the travel time a driver will make with the e-scooter. As a result, e-scooters offer drivers the opportunity to save time and manage their daily mobility more effectively, driving these vehicles attractive for transportation.

Keywords: E-scooter, time management, machine learning, prediction.

Öz

Bu çalışma, elektrikli scooter araçlarını tercih eden sürücülerin sürüş sürelerinin tahmin edilmesini amaçlamaktadır. E-scooter'lar genel olarak daha küçük boyutları ve manevra kabiliyetleri sayesinde şehir içi yolculuklarda hızlı ilerleme sağlayabildikleri için trafik sıkışıklığından kaynaklanan zaman kaybını azaltmaktadır. E-scooter'lar daha kompakt yapıları sayesinde park yeri bulma ve kolay park etme konusunda da avantaj sağlıyor. Bu çalışmada e-scooter araçlarını tercih eden sürücülerin sürüş sürelerinin tahmin edilmesi amacıyla ML algoritmaları kullanılmıştır. AB modeli, düşük Ortalama Kareler Hata (MSE) değeriyle (0,005) iyi performans gösterdi. Ortalama Karekök Hata (RMSE) ve Ortalama Mutlak Hata (MAE) değerleri de nispeten düşüktür (sırasıyla 0,069 ve 0,039), bu da modelin tahminlerinin gerçek değerlere yakın olduğunu göstermektedir. Ayrıca R-kare Belirleme Katsayısı (R2) değerinin (0,947) yüksek olması, bu modelin verileri oldukça iyi açıkladığını ve tahminlerinin gerçek değerlere yüksek doğrulukla yaklaştığını göstermektedir. Öte yandan GB algoritması, yüksek hata payı ve düşük doğruluk oranıyla farklı algoritmalara göre zayıf performans gösterdi. Bu sonuçlar, sürücünün e-scooter ile yapacağı yolculuk süresini tahmin ederek zaman yönetiminde avantaj sağlıyor. Sonuç olarak e-scooter'lar sürücülere zamandan tasarruf etme ve günlük hareketliliklerini daha etkin yönetme fırsatı sunarak bu araçları ulaşım açısından cazip hale getiriyor.

Anahtar Kelimeler: E-scooter, zaman yönetimi, makine öğrenimi, tahmin.

[1] PhD, Republic of Turkey Ministry of Transport and Infrastructure, Strategy Development Department, Head of Investment Management, hakan.inac@uab.gov.tr, Orcid: 0000-0001-9566-4106

Makale Türü: Araştırma Makalesi – Geliş Tarihi:06.02.2024 – Kabul Tarihi: 05.04.2024

DOI:10.17755/esosder.1432527

Atıf için: Elektronik Sosyal Bilimler Dergisi, 2024;23(91):1041-1057

1. INTRODUCTION

Electric scooters or e-scooters are a means of transportation that attracts attention, and usage is increasing nowadays, where urban transit is becoming increasingly popular (Ayözen et al., 2022). E-scooters are compact, two-wheeled transport vehicles powered by electric motors (Sun et al., 2021). Rechargeable lithium battery packs power these vehicles and move using electrical energy (Thackeray et al., 2012). Users control the speed by pressing the accelerator pedal. E-scooter vehicles usually range from 20 to 50 kilometers on a single charge (Capetillo & Berra, 2017). E-scooters offer an environmentally friendly alternative to fossil fuel vehicles (Flores & Jansson, 2021). These features make it especially attractive for environmentally conscious individuals as a sustainable and green transportation alternative.

E-scooters are ideal for urban transport (Gössling, 2020). E-scooters can move quickly and effectively in congested areas (Nocerino et al., 2016). The compact design of these vehicles makes parking easy, and e-scooters do not require large parking spaces like large vehicles. Therefore, e-scooters offer a practical solution to city traffic problems (Turon & Czech, 2022). E-scooters are a means of transportation that is extremely simple and accessible to drivers (Pazzini et al., 2022). These vehicles can usually be turned on or off with a single button, and speed control is done by pressing the pedal (Fishman & Cherry, 2016). It also offers many e-scooter rental services, allowing users to rent their scooter without owning it (Horton & Zeckhauser, 2016). E-scooters are becoming more and more popular around the world. E-scooters are mainly used in big cities and university campuses. Many companies promote this trend by offering shared electric scooter rental services (Inglesi-Lotz, 2016). In addition, many countries and cities have established regulations to regulate the use of electric scooters and ensure safety (Ignaccolo et al.,2022). As a result, e-scooters offer an environmentally friendly, fast, and practical alternative to urban transportation (Tuncer & Brown, 2020). Due to their easy-to-use design and global popularity, e-scooters will probably have a significant role in upcoming transportation systems.

Machine Learning (ML) is a subfield of artificial intelligence in which computer systems gain the ability to learn and gain experience based on data (Atalan et al., 2022; Atalan et al., 2023). This method allows algorithms to infer patterns and relationships from large data sets and use this information for future decisions and predictions (Schwendicke et al., 2020). ML can be used in many tasks, such as classification, regression, clustering, and prognosis, and its application areas are broad (İnaç et al., 2022). For example, ML is used in natural language processing, image recognition, medical diagnostics, autonomous driving in the automotive industry, financial analysis, marketing strategies, and many more (Sharma et al., 2021). ML is an essential part of the data science and artificial intelligence fields and has grown primarily due to advanced data processing technologies, powerful computing resources, and innovative algorithms (Akter et al., 2022). ML is a powerful tool for solving complex problems, extracting meaningful insights from data, and minimizing human intervention (Zhong et al., 2021). The effect and importance of ML technology are increasing daily with its use and development in more and more sectors (İnaç, 2023).

The link between ML and e-scooters plays a vital role in data-driven improvements in urban transport systems such as e-scooter rental and management (Vinagre et al., 2023). E-scooter sharing services, especially in big cities, generate data, and ML techniques offer great potential for analyzing and using this data (Ayözen et al., 2022; Zhao et al., 2022). By analyzing data such as the intensity of use of e-scooters in which regions there is more demand, with ML, companies can more effectively direct their services and plan the locations of e-scooters more intelligently (Haworth et al., 2021). In addition, ML is used in areas such as the safety of e-scooters and autonomous driving technologies

(Arslan and Uyulan, 2022; Fietz, 2020). Data collected through sensors and cameras can improve driver safety and help e-scooters automatically sense their surroundings and avoid hazards (Prabu et al., 2022). Therefore, the connection between ML and e-scooters is gaining importance as a collaboration that significantly improves transportation services and safety (Zuniga-Garcia et al., 2022).

This study aims to estimate the driving times of drivers who prefer e-scooter vehicles using ML algorithms. ML offers several methods that can be used to predict or analyze e-scooter ride times. These estimates can be used by e-scooter sharing services providers or city governments to assist in more efficient deployment, charging, and maintenance of e-scooters. ML is a versatile tool for analyses and predictions about e-scooter driving times (Das et al., 2023). The method to be used depends on the structure of the existing data set and the problem area. These estimates can provide valuable information to better plan and optimize e-scooter services.

In this study, estimation data of driving times of e-scooter drivers were obtained by using AdaBoost (AB), Random Forest (RF), and Gradient Boosting (GB) models. The algorithms of AB, RF, and GB are ensemble methods frequently used in machine learning (Atalan, 2023; Zhou et al., 2021; Li et al., 2019). The AB model combines weak learners (for example, decision trees with low success rates) to create a strong learner (Sagi and Rokach, 2018). Each stage weights the data by focusing on the errors of the previous step and trains the consecutive learners. The RF model is an ensemble method in which many decision trees (weak learners) are trained on a subset of random samples (Wan and Yang, 2013). Each tree is independent, and combining their results gives a more robust and generalizable model. The GB model is an ensemble method in which sequential learners (usually decision trees) are built on top of each other (Zhang et al., 2019). Each stage tries to correct the errors of the previous step and thus creates a stronger estimator. The main differences between these methods are based on factors such as training processes, weighting strategies, and how learners are brought together.

E-scooters have become a popular means of transportation in urban areas due to their compact size, manoeuvrability and environmental friendliness. This study aims to predict e-scooter users' driving times using machine learning (ML) algorithms, which is a new approach in the field of transportation research. The importance of this study is that it has practical implications for both e-scooter users and service providers. Using ML algorithms such as AdaBoost (AB), Random Forest (RF) and Gradient Boosting (GB), the study accurately predicts ride times based on factors such as rider age, gender, distance and e-scooter experience. Additionally, the study touches on the importance of ML in optimizing e-scooter services and infrastructure management. Moreover, this study demonstrates the potential of ML in improving the safety and autonomous capabilities of e-scooters. Overall, this research contributes to ongoing discussions on sustainable urban mobility and the integration of ML in transportation systems optimization.

This research comprises five primary sections. The initial section encompasses a review of the study's literature. Details regarding the research methodology are deliberated upon in the second segment. The numeric findings of the research are elucidated in the third portion. The study's subject, the employed methods, and its pros, cons, and limitations are addressed in the discussion section. Lastly, insights into the study's overall assessments and potential future contributions are examined within the concluding part of the research.

2. METHODOLOGY

The workflow that must be followed to develop an article using the obtained data includes a scientific method and a process that requires accuracy. The main steps of the operation of the technique developed for this study are expressed as follows:

The first step is to collect the raw data. This means managing or obtaining data related to the main subject of the research. In this study, data belonging to three input and one output variable were collected. The second step is cleaning and organizing the data and defining the variables. At this stage, it is essential to correct missing or incorrect values in the study data, manage the dataset, and bring it into a suitable format for analysis. The third step is to create the machine learning algorithms. It will be necessary to process the study data, extract features, and select the appropriate algorithm. This step involves setting, training, and tuning the correct algorithm and model. The fourth step includes determining the testing and training phases. Splitting study data into training and testing data is necessary to train ML models and evaluate their performance. The fifth step includes obtaining the performance measurement values of the algorithms. Different metrics are used to determine how well the model works. The criteria used for performance measurement values vary depending on the data type of the output variable. If the output variable has a categorical data type, measures such as accuracy, precision, recall, or ROC curve are used. However, the fact that the data of the output variable used in this study is numerical and continuous requires the calculation of RMSE, MSE, MAE, and R^2 values.

Finally, careful analysis should be performed in step six to confirm the validity of the results obtained. This step should be performed to ensure that the estimation results are compatible with the scope of the study and interpretation because the results must be presented in a way that contributes to the scientific community. The workflow of the proposed method for this study is shown in **Figure 1**.

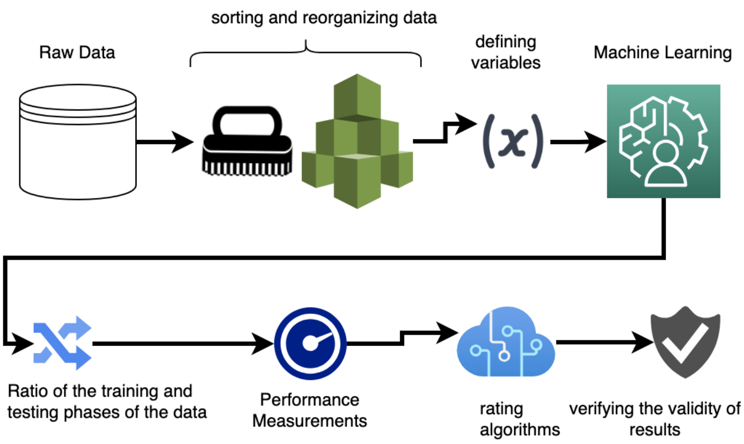

Figure 1. The flowchart of the methodology proposed

2.1. Data and Descriptive Statistics

In this study, driver age, gender, distance, and the number of drivers' driving experience with e-scooter vehicles were defined independently, while driving times were expressed as dependent variables. In this study, the driving times of 273 drivers were considered output variables. The data of the output variable are shown in **Figure 2**.

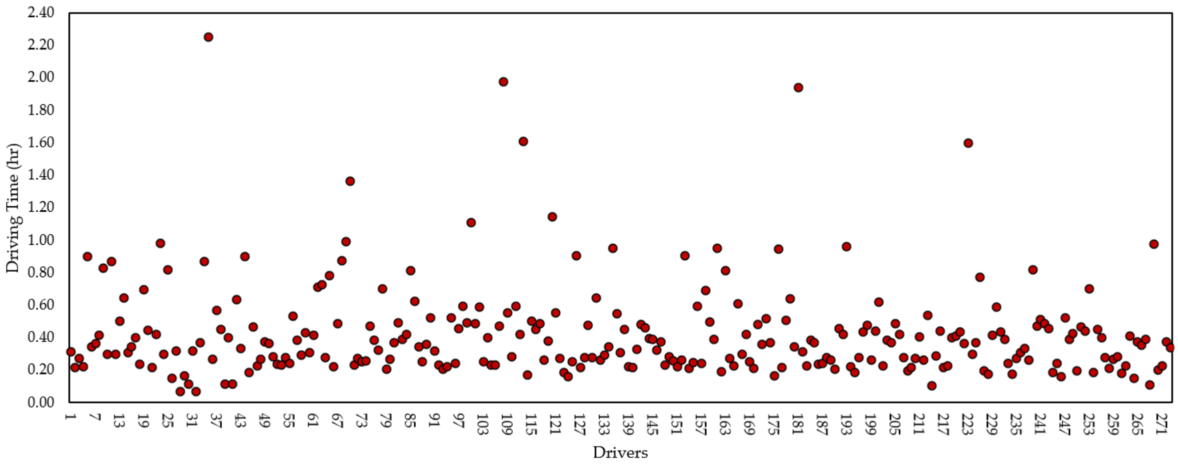

Figure 2. Time data of output variable

Descriptive statistics is a branch of statistics that includes statistical techniques and methods that summarize a data set and help us understand it. In this study, the values of the center, trend, and distribution measurements of the data were calculated. These statistics allow us to understand the central tendency of the data distribution, the extent of the distribution, and the shape of the data. Descriptive statistics are essential at the beginning of data analysis and interpretation processes. Measures of Central Tendency include values such as the mean, median, and mode of the dataset. The standard is calculated by dividing the sum of the data points by the number of data. The median refers to the middle value when ordering the data points from smallest to largest. Mode represents the most frequently occurring value in the dataset. Measures of Scatter show how dispersed the data points are. The standard deviation measures the spread of the data points around the mean. A more significant standard deviation indicates that the data points are more dispersed around the mean. The interquartile range (IQR) refers to the degree of spread of the middle half of the dataset.

This study calculated skewness and kurtosis values to evaluate whether the data distribution was symmetrical or skewed. A balanced distribution, such as the normal distribution, may make descriptive statistics easier to interpret, while a skewed distribution can make interpretation more complex. Descriptive statistics form a fundamental data analysis stage and play a critical role in understanding and explaining the data set. These statistics provide essential insights into the data and help prepare the data for further analysis. Descriptive statistics values of dependent and independent variables are shared in **Table 1** in detail.

Table 1. Descriptive statistics values of dependent and independent variables

Variable	Age		Tenure		Distance		Driving Time (hr)	
Gender	F	M	F	M	F	M	F	M
Total Count	66.00	207.0	66.00	207.0	66.00	207.0	66.00	207.0
Mean	24,87	22,13	32.14	72.99	2,367	2,090	0.555	0.383
StDev	7,621	6,519	53.19	107.1	3,601	2,510	0.368	0.251
Variance	58,07	42,49	2828	1148	12,96	6,298	0.135	0.063
Minimum	15,00	15,00	1.000	1.000	0.100	0.000	0.066	0.066
Q1	18,00	18,00	3.000	12.00	0.500	0.500	0.387	0.232
Median	25,00	18,00	14.50	31.00	1,000	1,200	0.438	0.290
Q3	30,00	25,00	35.75	100.0	2,750	2,900	0.604	0.467
Maximum	40,00	40,00	349.0	801.0	18,10	19,60	22,50	19,76
IQR	12,00	7,000	32.75	88.00	2,250	2,400	0.217	0.235
N for Mode	24.00	670.0	7.000	9.000	5.000	13.00	0.000	2.000
Skewness	0.780	1.000	3.890	3.670	3.070	3.060	2.760	2.710
Kurtosis	-0.340	0.670	19.60	18.60	10.12	14.40	9.320	10.68

2.2. Machine Learning (ML)

It is based on ML, data analysis, and model building, enabling computer systems to learn from experience. ML usually starts with the learning phase. At this stage, an algorithm uses large amounts of data to solve a particular task or problem. Data is often used to capture relationships between features and target results. The model training process is carried out through statistical analyses and mathematical calculations on this data. The model is adjusted to understand the complex relationships that develop based on the data and use it in future predictions. For this study, ML algorithms orange data mining program was used. The workflow of the preferred ML algorithms for obtaining forecast data is presented in **Figure 3**.

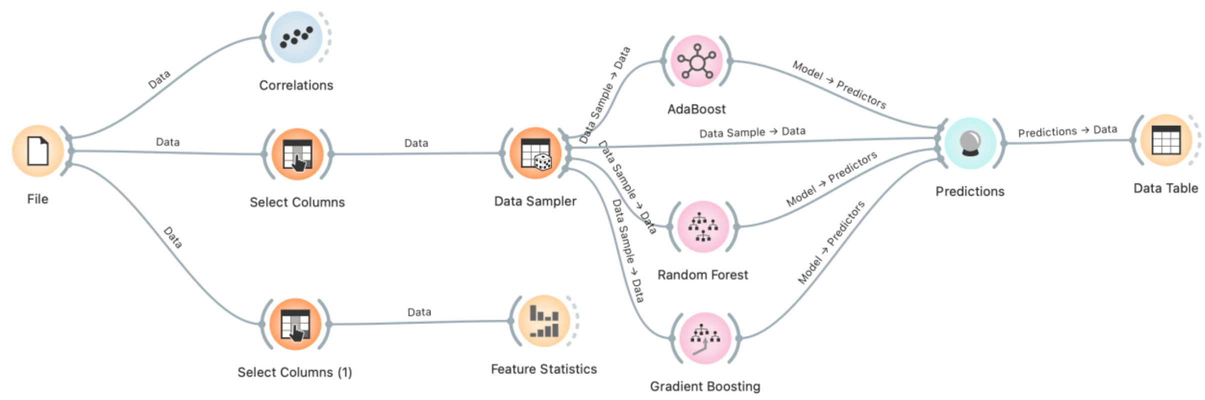

Figure 3. The workflow of the preferred ML algorithms

After model training, the ML algorithm is applied to new, unprecedented data. The model can make predictions or classifications based on new data by analyzing this data and using learned relationships. Therefore, ML is used in many applications, such as natural language processing, image recognition, automation, etc. These fundamental principles allow computers to solve complex problems by increasing their learning capabilities and play an essential role in artificial intelligence and data science.

The reliability of prediction data obtained by ML can vary depending on many factors. First, reliability depends on the quality and integrity of the dataset used. If the dataset on which the model is trained contains incomplete or incorrect information, the model's predictions may not be correct either. A clean and accurate dataset is critical to reliable forecasts. Second, reliability depends on the type and quality of the ML algorithm used. Some algorithms may perform better for specific data types or problems, while e-scooters may be less suitable for others. In addition, the correct tuning of the hyperparameters in the training process of the model also affects the reliability of the predictions. Also, constantly updating and retraining the model with new data can improve reliability. As a result, the reliability of prediction data obtained by ML depends on factors such as data quality, algorithm selection, and continuous monitoring and improvement of the model. Therefore, a careful data management and model development process is required to obtain reliable predictions.

In this study, AdaBoost or Adaptive Boosting (AB), Random Forest (RF), and Gradient Boosting (GB) models from ML algorithms were used. AB is an ensemble (mixed) learning algorithm used for classification problems. The basic idea is to combine low-performing classification models with weak learners to obtain a strong learner. AB gives weight to each weak learner and prioritizes examples that make more mistakes. Then, each weak learner is heavily retrained with a focus on reducing errors. This process creates a strong learner where weak learners come together. AB is a robust classification algorithm that gives successful results in many application areas, such as image recognition, text classification, and biomedical data analysis.

RF is a robust ensemble learning algorithm used for classification and regression problems. Its basic idea is to combine many decision trees to get a more powerful and stable model. Each tree is trained with random dataset samples, and a subset of the variables is used in this process. This allows trees to focus on different features and reduce the risk of overfitting. By combining the predictions of each tree, RF creates an ensemble model that is more reliable and can make better generalizable predictions. Therefore, it is a frequently preferred classification and regression tool for dealing with large datasets and complex problems.

GB is a robust ensemble learning algorithm used for classification and regression problems. Its basic principle is combining weak estimators (often used as decision trees) to create a stronger one. Initially, an estimate is made on the data, and an attempt is made to minimize the error between this estimate and the actual results. Next, the next forecast is made on this error, and this process continues until the forecast errors gradually decrease.

An essential feature of GB is that it uses a gradient descent method to reduce estimation errors at each step. This helps the model put more emphasis on training poor estimators and produce better predictions at each stage. GB is especially adequate for making high-accuracy predictions on large data sets and is used successfully in many application areas. Compared to other ensemble learning methods, such as AB and RF, GB offers more control and optimization possibilities and is often preferred in data science and machine learning.

2.3. Performance measurements of ML algorithms

Root Mean Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Error (MAE), and R-squared - Coefficient of Determination (R^2), which are used to evaluate the performance of ML algorithms, are widely used in the fields of ML and statistics to assess the performance of the prediction model. RMSE is used to measure the amount of error in a prediction model. It is calculated as the square root of the mean of the squares of the errors. This shows the magnitude of the differences between actual and predicted values. Higher RMSE values indicate that the model's predictions deviate more from the actual values. The RMSE is often used in regression analysis and provides a numerical measure of the amount of error. MSE is the square of the RMSE. It is the mean of the squares of the differences between actual and predicted values. This measures how erroneous the model's predictions are, but like the RMSE, it can emphasize significant errors more. MSE is also often used to evaluate the performance of regression models. MAE is the average of the absolute values of the differences between the actual values and the estimated values. This metric evaluates the magnitude of the amount of error but assumes that significant errors have the same weight as minor errors. Therefore, MAE is a more robust measure than RMSE and reduces the effect of outliers. R^2 is a statistical measure that measures the explanatory power of a regression model. Values are usually between 0 and 1. If the R^2 is close to 1, the model explains most of the observed variance and gives a perfect fit. However, if the R^2 is close to 0, the model does not explain the data and is not predictive. R^2 indicates how well the model fits the data, not how well it fits.

These four criteria are used to evaluate the performance of a prediction model and help make comparisons to determine which model produces better predictions. Each considers the success of the model from different aspects, so which metric to use depends on the requirements of the specific problem and the data structure.

3. RESULTS

In this study, regression analysis was performed to measure the effect of independent variables on dependent variables. Regression analysis is a statistical method used to examine the relationship between the dependent variable and one or more independent variables. The dependent variable is the primary variable that the analysis focuses on and that we are trying to predict. Independent variables are other variables that can affect or explain the dependent variable. Regression analysis aims to explain the relationship between these dependent and independent variables with a mathematical model. This model is determined according to the distribution and relationship of the data and expresses how the dependent variable depends on the independent variables and how strong or weak this relationship is. In this way, regression analysis is used to predict future events, understand cause-effect relationships, or examine interactions between variables. The ANOVA values of the independent variables in the regression analysis of the dependent variables in this study are given in **Table 2**.

Table 2. The ANOVA values of regression analysis

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.3697000	0.027000	13.670	0.001	
Age	-0.001061	0.000940	-1.130	0.260	1.03
Tenure	-0.000274	0.000066	-4.160	0.002	1.04
Distance	0.0931600	0.002280	40.940	0.001	1.01
Gender (M)	-0.137100	0.015300	-8.950	0.000	1.07

Coefficients, standard errors (SE Coef), t-values (T-Value), p-values (P-Value), and VIF (Variance Inflation Factor) values of the regression analysis of this study were calculated. The term constant represents its effect on the dependent variable. Generally, if the p value is less than 0.05, the independent variable has a statistical effect on the dependent variable (Nakagawa & Cuthill, 2007). In this case, the coefficient of the constant is 0.3697, and its standard error is 0.0270. The t-value was 13.67, and the p-value was 0.001. Since the p-value is very low, the constant term significantly affects the dependent variable. The coefficient of the age variable is -0.001061, and its standard error is 0.000940. The t-value was calculated as -1.13, and the p-value was 0.260. As the P-value was above

0.05 (a generally accepted alpha level), it was concluded that the age variable had no significant effect on the dependent variable. The coefficient of the service time variable is -0.000274, and its standard error is 0.000066. T-value was calculated as -4.16, and the p-value was calculated as 0.002. Since the p-value was very low, it was observed that the variable of service time had a significant effect on the dependent variable. The coefficient of the distance variable is 0.09316, and its standard error is 0.00228. T-value was 40.94, and the p-value was 0.001. Since the p-value is very low, the distance variable significantly affects the dependent variable. The gender variable has two categories: M (male) and another (default). The coefficient for males is -0.1371, and its standard error is 0.0153. T-value was calculated as -8.95 and the p-value as 0.000. Since the P-value is very low, it can be said that the gender variable has a significant effect on the dependent variable.

VIF (Variance Inflation Factor) measures the degree of multicollinearity between independent variables. If the VIF of each independent variable is greater than 1, there may be some degree of multicollinearity between these variables. In this study, all VIF values are between 1.01 and 1.07, indicating low multicollinearity. As a result, according to the results of the regression analysis given, it seems that the variables of constant term, length of service, distance, and gender significantly affect the dependent variable, but the effect of the age variable is not significant. Also, no multicollinearity problems are observed.

For this study, 80% and 20% separations were made for training and testing phases for running data in RF, AB, and GB algorithms from ML models. For each algorithm, 54 data were used for the test phase. The remaining data were used in the training phase for ML algorithms. The forecast data is compared with the actual data and visualized in **Figure 4**.

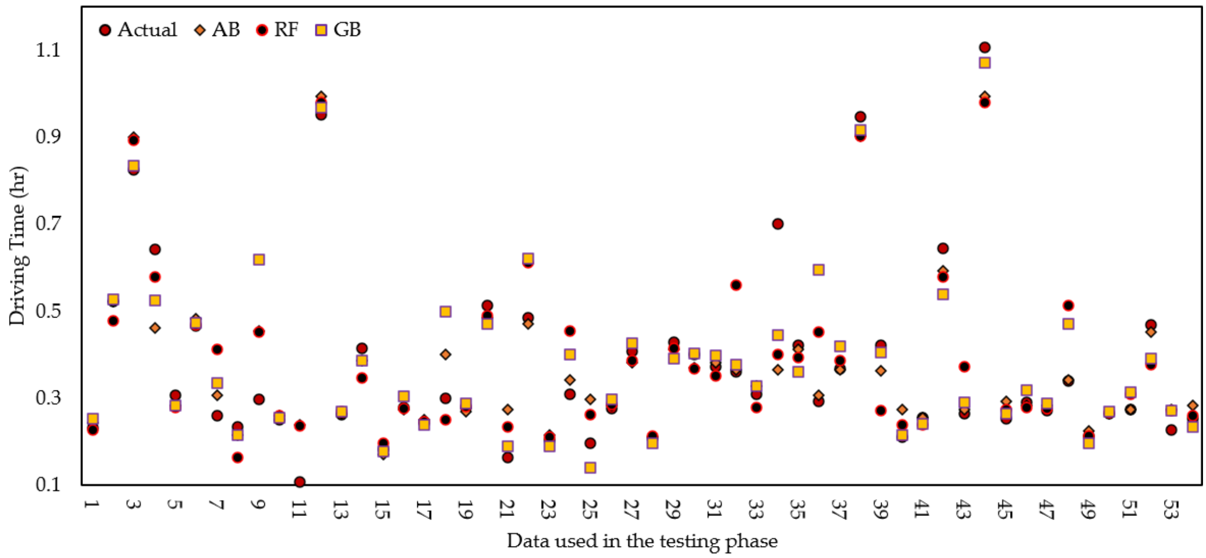

Figure 4: Prediction data based on ML models and actual data during the testing phase

The tree (tree) is used as the base estimator in the AB model. The model was created using a total of 50 estimators. The "Samme.r" algorithm was used because it is a classification model. However, "Linear" was chosen as a loss function used for regression. The training dataset has 219 data samples, and the characteristics of these samples include age (Age), gender (Gender), length of service (Tenure), and distance (Distance). The destination variable is named "Driving Time". This information describes the configuration of the model and the dataset used and indicates that the model is for classification purposes and uses the "Linear" loss function for regression operations.

The RF model used a tree assembly consisting of 10 trees, and each tree processed the data independently. The model does not limit the number of features considered, i.e., all features can be considered. However, the training process is not reproducible, meaning the results depend on randomly selected data samples and are not repeatable. Also, there are no limitations on the maximum tree depth and the minimum number of pieces to split a node. The dataset used for training has 219 data samples, and the characteristics of these samples include age (Age), gender (Gender), length of service (Tenure), and distance (Distance). The destination variable is named "Driving Time". This information describes the configuration of the model and the dataset used and demonstrates that the RF model makes predictions using independent trees and does not offer repeatable training.

The GB model has been implemented through the Scikit-learn library, with an explanation of the basic features and configuration of the GB model. The model uses a tree assembly of 100 trees, and each tree is trained to correct the errors of previous trees. The learning rate is set at 0.1, determining how much each tree contributes. The model is introduced with the repeatable training option, meaning the results can be repeated and compared. Also, the maximum tree depth was limited to 3, thus keeping the complexity of each tree under control. The dataset used for training has 219 data samples, and the characteristics of these samples include age (Age), gender (Gender), length of service (Tenure), and distance (Distance). The target variable (Target) is named "Driving Time." This information describes the basic configuration of the model and the dataset used, thus helping to better understand the model's performance and predictive capabilities.

Metrics such as RMSE, MSE, MAE, and R^2 are essential to evaluate the performance of ML algorithms. RMSE and MSE measure how much the estimates deviate from the actual values. Low RMSE and MSE values indicate that the model's predictions are close to the data and the model's accuracy is high. The MAE averages the absolute values of the prediction errors, which measures how close the model's predictions come to the data. R^2 determines how well the model explains the data; The closer it is to 1.00, the better the model. These metrics are used to compare the performance of different ML algorithms, optimize the settings of the model, and evaluate the generalization ability of the model. Therefore, these metrics are critical for model selection and evaluation of model quality and are used to understand how well the model works. The performance measurement values of the RF, AB, and GB algorithms used for this study were calculated. The results of four different metrics that measure the performance of ML models are shown in **Table 3**.

Table 3. Performance measurement values of ML models

Model	MSE	RMSE	MAE	R^2
AB	0.005	0.069	0.039	0.947
RF	0.007	0.086	0.052	0.917
GB	0.007	0.084	0.049	0.921

Table 3 presents comparisons over four different metrics that measure the performance of other ML models. The AB model performs well with a low MSE value (0.005). The RMSE and MAE values are also relatively low (0.069 and 0.039, respectively), indicating that the model's predictions are close to the actual values. Also, the high R^2 value (0.947) suggests that this model explains the data quite well, and its predictions approach the actual values with high accuracy.

The RF model performs lower than the other two, with a higher MSE value (0.007). The RMSE (0.086) and MAE (0.052) values are also higher than the other models, indicating that the model's predictions deviate more from the actual values. The R^2 value is 0.917, meaning that the RF model explains the data well but not as well as the AB model. The GB model performs similarly to the RF model, with a slightly lower MSE value (0.007) and RMSE value (0.084). The MAE value (0.049) is lower than the RF model. The R^2 value (0.921) indicates that the GB model explains the data well and that its predictions are correct.

As a result, according to these measurement values, the AB model performs better than the other two models. While the GB model also performs well, the RF model performs slightly less with higher error values. However, it is essential to consider other factors, such as problem context and data set, in model selection and interpretation of results.

Multidimensional Scaling (MDS) is used in machine learning and statistical data analysis. Its primary purpose is to transform high-dimensional data into a lower-dimensional space while preserving the similarity or distance between data points. This is useful in many applications, such as data visualization, analysis, classification, and clustering. MDS makes datasets containing many features or variables more understandable and interpretable. This method is used in many fields, such as data mining, marketing analysis, geographic information systems (GIS), and social sciences. Different variations of the MDS can be customized to the data and used to fit the problem context, making it a versatile analysis tool. The MDS image of this study is given in **Figure 5**.

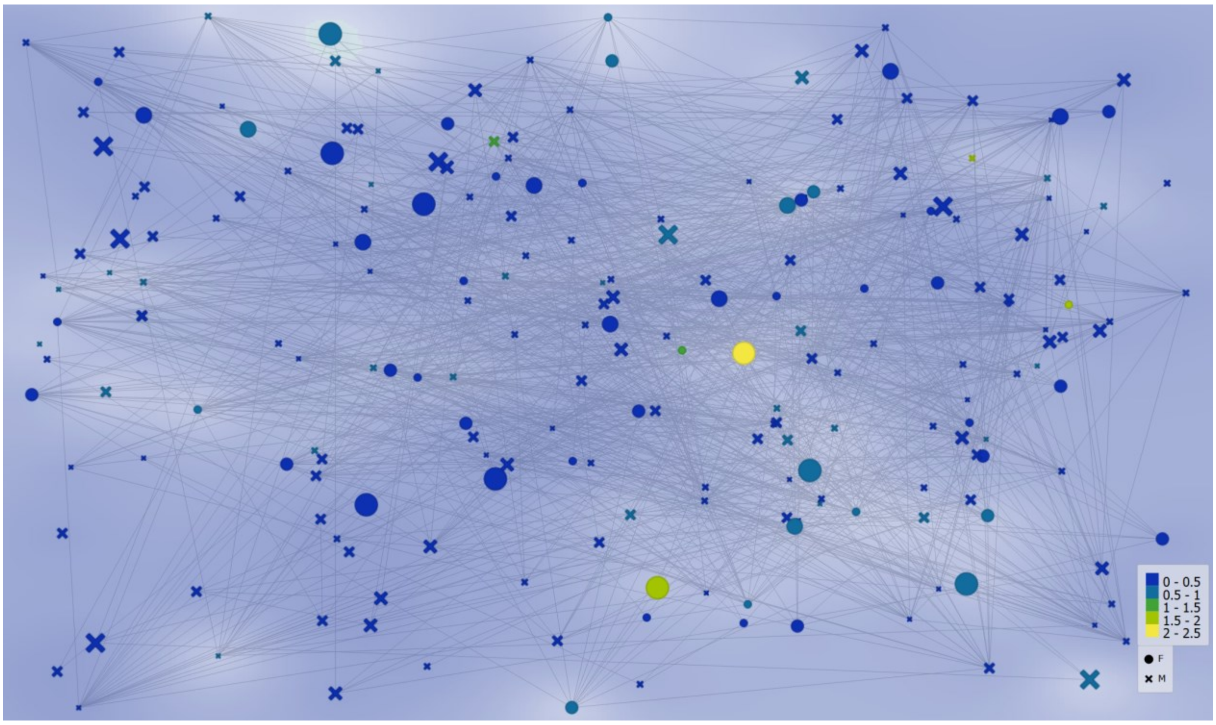

Figure 5. Multidimensional scaling of the ML algorithms

FreeViz is a diagram type used in machine learning and data visualization. This diagram helps us visually understand multidimensional data to understand better patterns, relationships, and groups in data sets. FreeViz is mainly used in data mining and analysis projects and is considered a powerful tool for interpreting and exploring data. This diagram represents data in multiple dimensions more simply and understandably and greatly helps analysts, decision-makers, and data scientists. FreeViz is essential for those who want to express data visually and makes data analysis processes more effective. The FreeViz diagram of this study is given in **Figure 6**. In this study, while tenure and gender are independent variables clustered together, age and driving distance affect the outcome variable without interaction.

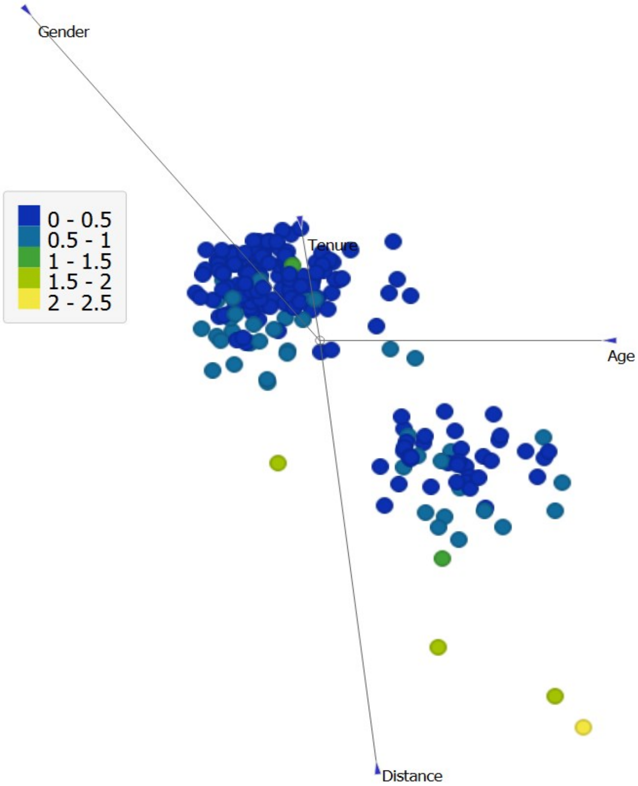

Figure 6. FreeViz diagram of the ML algorithms

4. DISCUSSION

E-scooters offer a more environmentally friendly, economical, and practical alternative to urban transportation (Nawaro, 2021). E-scooters are a more environmentally friendly transportation option than conventional internal combustion engine vehicles (Nocerino et al., 2020). These electrically powered vehicles provide zero-emission driving and contribute to reducing air pollution and greenhouse gas emissions. This improves air quality in cities and is essential in tackling climate change. E-scooters offer an effective solution to traffic jams. Increasing city traffic extends commuting time to work or traveling within the city. E-scooters provide a faster and more maneuverable alternative to circumventing traffic jams (Bretones et al., 2023).(Bretones et al., 2023)[15] Therefore, it makes urban transportation faster and more efficient.

E-scooters offer an economical and practical transportation option for users. Thanks to rental services, users can travel short distances quickly at affordable costs. E-scooters put an end to parking space problems because these vehicles can be parked easily thanks to their small size (James et al., 2019)(James et al., 2019)[16]. E-scooters can also be easily found and rented through mobile apps, making their means of transport quickly accessible (Tuncer and Brown, 2020). In conclusion, e-scooters are an environmentally friendly, economical, and practical transportation alternative that solves traffic problems. When used as part of sustainable city transport systems, it improves air quality, reduces congestion, and allows users to travel more efficiently. In facts, Personal e-scooter vehicles were widely preferred by the society during the pandemic periods when public transportation was prohibited (Atalan and Atalan, 2022). Therefore, the more widespread use of e-scooters in urban transport offers several advantages for individuals and the environment.

ML has advantages, disadvantages, and limitations in predicting e-scooter driving times (Moosavi et al. 2022). ML algorithms should be used to obtain predictive data because these algorithms can model the learning abilities of computers with the ability to identify complex patterns and relationships on large data sets. ML algorithms can capture more complex and non-linear relationships than traditional statistical methods. This is extremely valuable for discovering hidden patterns and interactions within the data, especially when working on large data sets. ML can process various data types and sizes, and this flexibility enables predictions in different domains. ML models can improve prediction accuracy (Park et al., 2021). The data analysis and modeling process can be optimized to highlight key features and connections in the data to make accurate predictions (Atalan et al., 2018). This helps businesses, financial institutions, healthcare, and many other industries better predict future events (Dönmez and Atalan, 2019). ML supports making decisions based on large amounts of data and optimizing business processes. Especially in recommendation systems and personalized advertising, these algorithms can make recommendations based on data to understand users' interests and preferences. In addition, ML adds value to data-driven decision-making processes in many areas, such as financial risk analysis, health diagnosis, and logistics management.

Studies on the prediction of driving times of electric scooters attract significant attention with the use of ML methods. Ayözen et al. successfully applied machine learning algorithms to predict driving times of electric scooters (Ayözen et al., 2022). In this study, different machine learning techniques such as AdaBoost, Random Forest and Gradient Boosting were used to accurately predict driving times. The results obtained in the study show that high levels of accuracy are achieved in estimating driving times. Teusch et al. ML methods were used to estimate the driving times of electric scooters. This study has provided valuable insights to help drivers manage their daily mobility more effectively (Teusch et al., 2023). Chen et al. successfully used machine learning techniques such as Gradient Boosting to predict electric scooter ride times. This study has provided significant progress in terms of accuracy and efficiency in estimating driving times (Chen et al., 2017).

ML algorithms are powerful for working on large and complex data sets, increasing prediction accuracy, and making data-driven decisions (Schwendicke et al., 2020). For this reason, many industries and businesses prefer ML for data analysis and prediction purposes, and using these algorithms is becoming increasingly common (Schwendicke et al., 2020; Boukerche and Wang, 2020; Khan et al., 2020). ML can use large datasets to predict e-scooter ride times. This makes it possible to make more precise predictions by analyzing user behavior and influencing factors (weather conditions, traffic density, time, starting and destination points, etc.) (Peng et al., 2022; Schwendicke et al., 2020; Fuentas et al., 2020; Li et al., 2018). ML models can adapt to changing needs over time. Appropriate algorithms can estimate driving times based on seasons, holidays, or special events. E-scooter service providers can better plan the deployment and recharge times of e-scooters with ML-predicted driving times. This can reduce operational costs and improve service quality.

Data quality can affect the accuracy of estimates. Missing or incorrect data can cause prediction errors and adversely affect model performance. Updating and training ML models takes time and resources. Constantly updating the model can be an operational challenge. People's driving behavior can be complex and difficult to predict. Any prediction model may not fully capture individual preferences and variable human behavior. The privacy of the personal data of e-scooter users is essential, and the processing and storage of this data may be subject to regulatory requirements. All in all, ML can be a powerful tool for estimating e-scooter driving times. Still, it may face some challenges, such as data quality, model updates, and the complexity of human behavior. With the correct data and an appropriate model, e-scooter service providers can make their operations more efficient and improve the user experience. However, managing these processes and data security should be handled with care.

5. CONCLUSION

The relationship between ML and e-scooter driving time can be understood using data analysis and prediction models. E-scooter driving time is considered a complex variable influenced by several factors. ML is used to identify and predict the contribution of these factors to driving time. First, ML models do data mining and data analysis to identify critical factors affecting driving time. These factors can be various variables, such as starting and destination points, road length, weather conditions, traffic density, and time of day. ML can quantify the impact of these factors on driving time. Appropriate ML algorithms, such as regression analysis or time series models, are then used to capture and predict the effects of these factors. These models can predict driving time by identifying patterns within the data. For example, a regression model can estimate typical driving time for a given starting and destination point. In conclusion, ML is a powerful tool for identifying and predicting factors that affect e-scooter driving time. This can help e-scooter service providers deliver better service, allowing users to plan and travel more efficiently.

E-scooters have the potential to become an essential means of transport in the future, and this potential is due to many reasons. Because e-scooters run on electricity, these vehicles offer a more environmentally friendly transportation option than conventional internal combustion engines. Reducing the use of fossil fuels improves air quality and reduces greenhouse gas emissions. Traffic congestion in cities is a serious problem. E-scooters provide a fast and efficient transportation option over short distances, which can reduce traffic congestion. This can help cities become more efficient and sustainable. E-scooters are a means of transportation with a low cost of ownership and limited cost of use. Rent or share models allow people to use e-scooters at an affordable price. E-scooters can contribute to the creation of sustainable cities. These vehicles can be essential for urban planners who want to design greener and more accessible cities.

E-scooters are ideal for covering short distances quickly. These vehicles can become a preferred means of transportation for city meetings, shopping trips, or leisure trips. E-scooters increase transportation options and give people more mobility. E-scooters can be an essential alternative, especially for individuals with physical disabilities or limited mobility. If e-scooters become available in more cities, the social, environmental, and economic benefits of this mode of transportation will become even more significant. However, appropriate policy and infrastructure must be established to regulate use and management.

REFERENCES

Akter, S., Michael, K., Uddin, M. R., McCarthy, G., & Rahman, M. (2022). Transforming business using digital innovations: The application of AI, blockchain, cloud and data analytics. *Annals of Operations Research*, 1-33.

Arslan, E., & Uyulan, Ç. (2023). Analysis of an e-scooter and rider system dynamic response to curb traversing through physics-informed machine learning methods. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, *237*(7), 1555-1571.

Atalan, A. (2023, May). Neural network and random forest algorithms for estimation of the waiting times based on the DES in ED. *In International Conference on Contemporary Academic Research,1*(1), 14-20.

Atalan, A. (2023). Forecasting drinking milk price based on economic, social, and environmental factors using machine learning algorithms. *Agribusiness*, *39*(1), 214-241.

Atalan, A., & Atalan, Y. A. (2022). Analysis of the impact of air transportation on the spread of the covid-19 pandemic. *In Challenges and Opportunities for Transportation Services in the Post-COVID-19 Era* (pp. 68-87). IGI Global.

Atalan, A., Dönmez, C. Ç., & Atalan, Y. A. (2018). Yüksek-eğitimli uzman hemşire istihdamı ile acil servis kalitesinin yükseltilmesi için simülasyon uygulaması: Türkiye sağlık sistemi. *Marmara Fen Bilimleri Dergisi*, *30*(4), 318-338.

Atalan, A., Şahin, H., & Atalan, Y. A. (2022, September). Integration of machine learning algorithms and discrete-event simulation for the cost of healthcare resources. In *Healthcare* (Vol. 10, No. 10, p. 1920). MDPI.

Ayözen, Y. E., İnaç, H., Atalan, A., & Dönmez, C. Ç. (2022). E-Scooter micro-mobility application for postal service: the case of turkey for energy, environment, and economy perspectives. *Energies*, *15*(20), 7587.

Boukerche, A., & Wang, J. (2020). Machine learning-based traffic prediction models for intelligent transportation systems. *Computer Networks*, *181*, 107530.

Bretones, A., Marquet, O., Daher, C., Hidalgo, L., Nieuwenhuijsen, M., Miralles-Guasch, C., & Mueller, N. (2023). Public health-led insights on electric micro-mobility adoption and use: a scoping review. *Journal of Urban Health*, 1-15.

Capetillo, A., & Ibarra, F. (2017). Multiphase injector modelling for automotive SCR systems: A full factorial design of experiment and optimization. *Computers & Mathematics with Applications*, *74*(1), 188-200.Flores, P. J., & Jansson, J. (2021). The role of consumer innovativeness and green perceptions on green innovation use: The case of shared e-bikes and e-scooters. *Journal of Consumer Behaviour*, *20*(6), 1466-1479.

Chen, X. M., Zahiri, M., & Zhang, S. (2017). Understanding ridesplitting behavior of on-demand ride services: An ensemble learning approach. *Transportation Research Part C: Emerging Technologies*, *76*, 51–70.

Das, S., Hossain, A., Rahman, M. A., Sheykhfard, A., & Kutela, B. (2023). Case study on the traffic collision patterns of e-scooter riders. *Transportation Research Record*, 03611981231185770.

Dönmez, C. Ç., & Atalan, A. (2019). Developing statistical optimization models for urban competitiveness index: under the boundaries of econophysics approach. *Complexity*, *2019*, 1-11.

Fietz, L. E. (2020). *Predicting hourly shared e-scooter use in chicago: a machine learning approach* (Doctoral dissertation, University of Oregon).

Fishman, E., & Cherry, C. (2016). E-bikes in the Mainstream: Reviewing a Decade of Research. *Transport reviews*, *36*(1), 72-91.

Fuentes, S., Gonzalez Viejo, C., Cullen, B., Tongson, E., Chauhan, S. S., & Dunshea, F. R. (2020). Artificial intelligence applied to a robotic dairy farm to model milk productivity and quality based on cow data and daily environmental parameters. *Sensors*, *20*(10), 2975.

Gössling, S. (2020). Integrating e-scooters in urban transportation: Problems, policies, and the prospect of system change. *Transportation Research Part D: Transport and Environment*, *79*, 102230.

Haworth, N., Schramm, A., & Twisk, D. (2021). Comparing the risky behaviors of shared and private e-scooter and bicycle riders in downtown Brisbane, Australia. *Accident Analysis & Prevention*, *152*, 105981.

Horton, J. J., & Zeckhauser, R. J. (2016). Owning, using and renting: some simple economics of the" sharing economy". *National Bureau of Economic Research.* (No. w22029).

Ignaccolo, M., Inturri, G., Cocuzza, E., Giuffrida, N., Le Pira, M., & Torrisi, V. (2022). Developing micromobility in urban areas: network planning criteria for e-scooters and electric micromobility devices. *Transportation research procedia*, *60*, 448-455.

İnaç, H. (2023). Micro-mobility sharing system accident case analysis by statistical machine learning algorithms. *Sustainability*, *15*(3), 2097.

İnaç, H., Ayözen, Y. E., Atalan, A., & Dönmez, C. Ç. (2022). Estimation of postal service delivery time and energy cost with e-scooter by machine learning algorithms. *Applied Sciences*, *12*(23), 12266.

Inglesi-Lotz, R. (2016). The impact of renewable energy consumption to economic growth: A panel data application. *Energy economics*, *53*, 58-63.

James, O., Swiderski, J. I., Hicks, J., Teoman, D., & Buehler, R. (2019). Pedestrians and e-scooters: An initial look at e-scooter parking and perceptions by riders and non-riders. *Sustainability*, *11*(20), 5591.

Khan, P. W., Byun, Y. C., Lee, S. J., Kang, D. H., Kang, J. Y., & Park, H. S. (2020). Machine learning-based approach to predict energy consumption of renewable and nonrenewable power sources. *Energies*, *13*(18), 4870.

Li, K., Zhou, G., Zhai, J., Li, F., & Shao, M. (2019). Improved PSO_AdaBoost ensemble algorithm for imbalanced data. *Sensors*, *19*(6), 1476.

Li, Y., Zou, C., Berecibar, M., Nanini-Maury, E., Chan, J. C. W., Van den Bossche, P., ... & Omar, N. (2018). Random forest regression for online capacity estimation of lithium-ion batteries. *Applied energy*, *232*, 197-210.

Moosavi, S. M. H., Ma, Z., Armaghani, D. J., Aghaabbasi, M., Ganggayah, M. D., Wah, Y. C., & Ulrikh, D. V. (2022). Understanding and predicting the usage of shared electric scooter services on university campuses. *Applied Sciences*, *12*(18), 9392.

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, *82*(4), 591–605. https://doi.org/10.1111/j.1469-185X.2007.00027.x

Nawaro, Ł. (2021). E-scooters: competition with shared bicycles and relationship to public transport. *International Journal of Urban Sustainable Development*, *13*(3), 614-630.

Nocerino, R., Colorni, A., Lia, F., & Lue, A. (2016). E-bikes and E-scooters for smart logistics: environmental and economic sustainability in pro-E-bike Italian pilots. *Transportation research procedia*, *14*, 2362-2371.

Park, J. G., Jun, H. B., & Heo, T. Y. (2021). Retraining prior state performances of anaerobic digestion improves prediction accuracy of methane yield in various machine learning models. *Applied Energy*, *298*, 117250.

Pazzini, M., Cameli, L., Lantieri, C., Vignali, V., Dondi, G., & Jonsson, T. (2022). New micromobility means of transport: An analysis of e-scooter users' behaviour in Trondheim. *International journal of environmental research and public health*, *19*(12), 7374.

Peng, H., Nishiyama, Y., & Sezaki, K. (2022). Assessing environmental benefits from shared micromobility systems using machine learning algorithms and Monte Carlo simulation. *Sustainable Cities and Society*, *87*, 104207.

Prabu, A., Shen, D., Tian, R., Chien, S., Li, L., Chen, Y., & Sherony, R. (2022). A wearable data collection system for studying micro-level e-scooter behavior in naturalistic road environment. *arXiv preprint arXiv:2212.11979*.

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1249.

Schwendicke, F. A., Samek, W., & Krois, J. (2020). Artificial intelligence in dentistry: chances and challenges. *Journal of dental research*, *99*(7), 769-774.

Sharma, N., Sharma, R., & Jindal, N. (2021). Machine learning and deep learning applications-a vision. *Global Transitions Proceedings*, *2*(1), 24-28.

Sun, B., Garikapati, V., Wilson, A., & Duvall, A. (2021). Estimating energy bounds for adoption of shared micromobility. *Transportation Research Part D: Transport and Environment*, *100*, 103012.

Teusch, J., Gremmel, J. N., Koetsier, C., Johora, F. T., Sester, M., Woisetschläger, D. M., & Müller, J. P. (2023). A Systematic Literature Review on Machine Learning in Shared Mobility. *IEEE Open Journal of Intelligent Transportation Systems*, *4*, 870–899.

Thackeray, M. M., Wolverton, C., & Isaacs, E. D. (2012). Electrical energy storage for transportation—approaching the limits of, and going beyond, lithium-ion batteries. *Energy & Environmental Science*, *5*(7), 7854-7863.

Tuncer, S., & Brown, B. (2020, April). E-scooters on the ground: Lessons for redesigning urban micro-mobility. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-14).

Turoń, K., & Czech, P. (2020). The concept of rules and recommendations for riding shared and private e-scooters in the road network in the light of global problems. In *Modern Traffic Engineering in the System Approach to the Development of Traffic Networks: 16th Scientific and Technical Conference" Transport Systems. Theory and Practice 2019" Selected Papers 16* (pp. 275-284). Springer International Publishing.

Vinagre Díaz, J. J., Fernández Pozo, R., Rodríguez González, A. B., Wilby, M. R., & Anvari, B. (2023). Blind classification of e-scooter trips according to their relationship with public transport. *Transportation*, 1-22.

Wan, S., & Yang, H. (2013, July). Comparison among methods of ensemble learning. In *2013 International Symposium on Biometrics and Security Technologies* (pp. 286-290). IEEE.

Zhang, C., Zhang, Y., Shi, X., Almpanidis, G., Fan, G., & Shen, X. (2019). On incremental learning for gradient boosting decision trees. *Neural Processing Letters*, *50*, 957-987.

Zhao, P., Li, A., Pilesjö, P., & Mansourian, A. (2022). A machine learning based approach for predicting usage efficiency of shared e-scooters using vehicle availability data. *AGILE: GIScience Series*, *3*, 20.

Zhong, S., Zhang, K., Bagheri, M., Burken, J. G., Gu, A., Li, B., ... & Zhang, H. (2021). Machine learning: new ideas and tools in environmental science and engineering. *Environmental Science & Technology*, *55*(19), 12741-12754.

Zhou, X., Tian, S., An, J., Yang, J., Zhou, Y., Yan, D., ... & Jin, X. (2021). Comparison of different machine learning algorithms for predicting air-conditioning operating behavior in open-plan offices. *Energy and Buildings*, *251*, 111347.

Zuniga-Garcia, N., Tec, M., Scott, J. G., & Machemehl, R. B. (2022). Evaluation of e-scooters as transit last-mile solution. *Transportation research part C: emerging technologies*, *139*, 103660.