




## Article

# An Application to Predict Range of Electric Two-Wheeler Using Machine Learning Techniques

Al Amin <sup>1</sup>, Mohammad Shafenoor Amin <sup>1</sup> and Chongdu Cho <sup>2,\*</sup>

<sup>1</sup> Program in Future Vehicle Engineering, Inha University, 100 Inha-ro, Michuhol-gu, Incheon 22212, Republic of Korea; alamin@inha.edu (A.A.); s.amin@inha.edu (M.S.A.)

<sup>2</sup> Department of Mechanical Engineering, Inha University, 100 Inha-ro, Michuhol-gu, Incheon 22212, Republic of Korea

\* Correspondence: cdcho@inha.ac.kr; Tel.: +82-32-860-7321

**Abstract:** Electric two-wheelers are becoming increasingly popular across the world, particularly in cities where their small size and flexibility make them a viable option for navigating congested streets. One of the most challenging aspects of e-mobility on two-wheelers is precisely calculating their range. This might be an issue for riders who must go long distances or who have limited access to charging stations. Various factors can influence an electric two-wheeler range, making it challenging to predict how far it can travel on a single charge. To tackle this problem, most of the manufacturers offer range predictions based on both test data and real-world usage scenarios. However, these estimates are customized for specific vehicle models and testing parameters that may not apply in all circumstances. Additionally, it can be challenging to obtain comprehensive technical specifications for two-wheelers available in the market, as most manufacturers do not provide detailed technical information. Hence, it is crucial to address the challenge of range prediction for two-wheelers in general, which can be advantageous for riders. In this paper, we discuss the precise prediction of the remaining range of electric two-wheelers even without knowing detailed e-mobility technical information. An application is also developed only for this research purpose, which can provide navigation services. Our approach concentrates on user behavior, weather, road conditions, and the vehicle's performance history, which is gathered through the application. The collected data are used to train the selected ML model on the cloud. We applied various machine learning algorithms before deploying in the cloud where the SVM algorithm demonstrated outstanding performance, with a mean absolute error of 150 m for an average distance of 7.46 km. Furthermore, the model's performance was evaluated after deployment and tested having 130 m error on average.

**Keywords:** e-mobility; electric two-wheeler; range prediction; machine learning algorithms



**Citation:** Amin, A.; Amin, M.S.; Cho, C. An Application to Predict Range of Electric Two-Wheeler Using Machine Learning Techniques. *Appl. Sci.* **2023**, *13*, 5840. <https://doi.org/10.3390/app13105840>

Academic Editor: Jose Machado

Received: 28 March 2023

Revised: 26 April 2023

Accepted: 6 May 2023

Published: 9 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

An electric vehicle (EV) is a type of vehicle that is powered entirely by electricity, as opposed to vehicles that are powered by fossil fuels. EVs are becoming an increasingly popular alternative to traditional gasoline-powered vehicles due to their environmental benefits and lower operating costs. Gasoline-powered vehicles emit harmful pollutants into the air, contributing to air pollution and climate change. Meanwhile, electric vehicles do not produce any emissions during operation, making them a much more sustainable and eco-friendly option [1,2]. They also require less maintenance and can be cheaper to operate over the long term. Electric vehicles refer to any type of vehicle that is powered by an electric motor or battery, rather than relying solely on a traditional combustion engine. This includes various types of vehicles, such as two-wheelers, three-wheelers, and four-wheelers.

Despite the many benefits of reduced emissions and lower operating costs that four-wheelers provide, there are still obstacles that need to be overcome to make them a feasible

choice for all drivers, including challenges such as range anxiety [3,4], a lack of charging infrastructure, high upfront costs, and long charging times. Similarly, two-wheeler e-mobility (such as e-bikes and scooters) also has its advantages, but it shares the same concerns as four-wheelers, including range anxiety, which is a common concern among electric two-wheeler users who worry about running out of battery power while on the road [5]. Even though e-bikes can alleviate the riders' concerns about range anxiety by providing pedals, which can assist in conserving battery power, it is still necessary to determine the exact range.

The main difference between range prediction for four-wheelers and two-wheelers is the size of the battery and the energy consumption of the vehicle where the estimation of range is primarily dependent on the battery of the vehicle. Electric four-wheelers usually have larger battery packs, which means they can travel longer distances on a single charge compared to two-wheelers [6]. As a result, range prediction for four-wheelers usually involves more complex models. On the other hand, two-wheelers have smaller battery packs and are typically used for shorter trips, such as commuting or running errands. Range prediction for electric two-wheelers is, therefore, simpler and can often be done using basic information, such as the initial battery charge, bike and engine type, and the weight of the rider. Another key difference is the impact of speed on range prediction. Four-wheelers are typically designed to travel at higher speeds, which means they consume more energy, and their range is more affected by factors such as wind resistance and drag. Two-wheelers, on the other hand, have lower maximum speeds and are less affected by these factors. This means that range prediction for two-wheelers is less sensitive to variations in speed compared to four-wheelers.

Electric two-wheelers typically come with a digital display, which indicates battery voltage. However, this does not help the consumers, as they cannot be translated to the remaining range of the vehicle. The battery size and current state of charge determine how much further the rider can go. Therefore, a more precise calculation might increase the likelihood of e-bike adoption even more [4]. Some e-mobility provides range information based on assistance level required from the drive unit [7]. However, there are other factors, such as battery size and type, vehicle weight, terrain, weather conditions, and rider behavior, that play an important role in prediction of range other than the drive unit [8]. Therefore, the accuracy of the information is not sufficient for the current riders. This creates anxiety among e-mobility owners about inadequate battery charges [9]. It is possible to obtain this information using several sensors, as modern e-bikes include sensors that transmit data to the control unit, such as speed, cadence, torque, and battery temperature. Moreover, this information can be transmitted to a smartphone for further use [10]. However, some of them need Bluetooth connection and extra hardware setup, which is a hassle for the rider. There are some applications where Bluetooth connection is not required but, in those cases, to provide remaining range or power, riders need to give a lot of inputs regarding the technical details of their two-wheelers [11].

In this paper, we have considered all these factors and estimated the range for two-wheelers by using vehicle information, rider behavior, road, and weather condition. We have developed a data-driven machine-learning-based strategy and an application was also developed to obtain all required data inspired by [12]. A very limited number of works from the past have dealt with the same issue. Our goal is to predict range based on the remaining battery life using machine learning and provide desired information to riders without putting in much effort and/or technical information regarding e-mobility. In the next section, we will conduct an analysis of the literature that we believe is pertinent to the work that we are conducting.

## 2. Related Works

There have been numerous studies conducted on range prediction for electric vehicles; most of these studies have focused on three- or four-wheeler electric vehicles. However, there are only a few studies that have specifically addressed range prediction for electric

bikes or two-wheelers. This highlights the need for further research in this area to improve the accuracy of range prediction for electric two-wheelers.

Range prediction is a standard feature when it comes to four-wheeled electric vehicles (EV) to inform the drivers when to charge the vehicle. To overcome the range prediction problem in the case of EV, physical models using real-time data and artificial intelligence models both trained on previous vehicle data have been investigated to estimate the range of EV where physical models are based on their ability to provide an estimate of energy consumption by modeling either the battery pack or the vehicle itself (see, e.g., [13–21]). However, when it comes to the two-wheeler e-mobility especially, there are very few works that have been conducted [8,10–12,22–24].

We present a brief overview of the related research we have compiled in Table 1 where we mentioned details of the e-mobility type and approach of each study. There are some studies that focused on battery power consumption rather than range estimation. We know the main power source of a two-wheeler is the battery, which is one of the factors to determine the range. In [22], the authors studied battery power and developed a mathematical model for power consumption based on road slope, friction, air resistance, rolling resistance, user information, and engine type, along with bike characteristics. An application was also developed to give input parameters in the mathematical model and to check each segment's battery consumption in a planned route. Two e-bikes were prepared to test their model and found that the prediction forecast is closer to the actual battery consumption. We found another similar study about battery consumption [10] where artificial neural networks were used to calculate rough estimates of travel time and battery consumption along a given route and then presented through visualization to the user. Evolutionary algorithms used these approximations as part of their optimization processes. The outcomes for trips taken with and without the optimization system have been compared. However, mean absolute error was found to be around 9.7% without optimization and a much better result was achieved after optimization. Moreover, the result was not the same for two similar batteries.

**Table 1.** Related work summary.

Ref.	E-Mobility Type	Approach
[22]	E-bike	Mathematical approach based on battery consumption data to predict power consumption
[10]	E-bike	ANN approach based on vehicle, smartphone sensor, and geographic information to predict power consumption
[11]	Two-wheeler	Simulink model approach based on vehicle and battery information to predict mileage
[8]	E-bike	Machine learning approach based on SOC, rider, and environment information to predict destination can be reached with current SOC or not
[12]	E-bike	Machine learning approach based on rider, traffic, and battery information to predict battery consumption
[23]	Scooter	Simulink model approach based on rider, vehicle, and battery information to predict mileage
[24]	Scooter	Simulink model approach based on rider, vehicle, and battery information to predict power and energy

We understand that battery consumption is directly related to range prediction. However, our focus is on estimating the maximum distance that can be covered by the vehicle before the battery runs out of power. In [11], they ensured the safe and effective operation of e-mobility systems by better understanding how batteries react to a wide range of conditions through the development of a battery simulation model. A two-wheeler was prepared for on-road test in Italy to measure the battery information during each test ride. Then, the test data were examined through their models, which are battery model, longitudinal dynamic model, and global model, where all the models were studied through Matlab Simulink. Battery models have been prepared using battery information such as voltage, current, total energy, discharge current, weight, and so on. In the dynamic model, it is

based on the kinematics of the vehicle with the velocity profile over time and, finally, in the global model, authors connected the battery model and dynamic model together. It seems that the global model carried out better results than other models. However, these experiments needed lots of input parameters, such as voltage, capacity, discharge current, battery weight, and many more, where riders may not know all the electronics data.

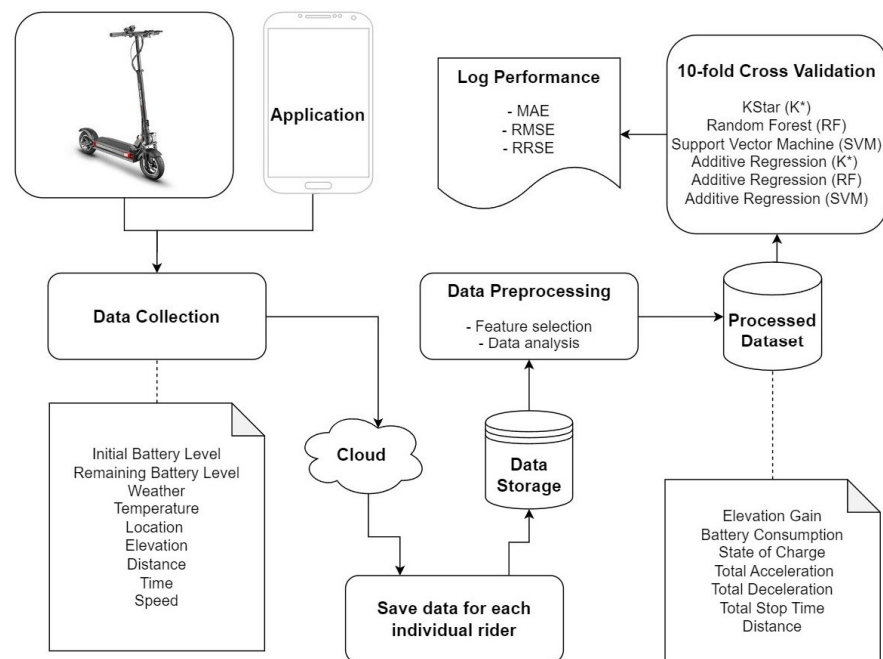
In [23], the authors attempted a comparable model, resulting in an error range of approximately 0.36% to 1.1% in predicting range. Another similar model was also developed by authors in [24], where an electric scooter was studied to predict energy consumption and range through dynamometer test and simulation model. The simulation model shows that the electric scooter can cover a distance of up to 63,470 m. During the dynamometer tests, the scooter achieved an average range of 60,460 m, while, on normal on-road testing, it covered 62,830 m. The errors in comparison are 4.74% and 1.01% for the standard dynamometer and on-road tests, respectively. After reviewing simulation-based study, we shifted our focus to machine-learning-based study. We have discovered in [12] a machine learning model was used based on the battery information and sensor data from the e-bike. The authors from the University of Waterloo developed a project called WeBike, where an e-bike was prepared with 31 sensors and connected with an online map server through a GPS-equipped smartphone to collect data. They have organized several participants to collect data such as each participant's battery consumption rate, average speed, trip segments, traffic sign, average battery temperature, etc. Two prediction models such as mean prediction and linear regression are used for range prediction that took into account the behavior of the cyclist in addition to the route. Both model's root mean squared errors were around 5.3~9.7%. Moreover, two-wheelers do not come with multiple sensors unless someone upgrades by himself/herself like they prepared their own hardware set. We focused on data which can be easily accessible without any hassle from the rider and without any Bluetooth connection. The findings were good enough but there are other external factors as well, such as actual battery capacity, rider's speed, wind speed, and elevation, which also need to be studied, explained by the author.

In another study [8], the authors worked with those external factors where they developed a system that can be installed in a smartphone. For the study purpose, a prototype e-bike was prepared where their motor control system, microcontroller, and battery were highlighted. The MCS app which they have developed can read battery information through Bluetooth as well as collect smartphone sensor data. The data from bike dependency, bicyclist dependency (bike effort and previous driving profile), and environment dependency (temperature, wind, etc.) are combined to make the prediction. They have experimented with Naïve Bayes algorithm to predict whether the user can go for the trip with current battery charge or not and the approximate charge it may take to complete the trip. Still, it was claimed that the approach needs to improve due to weather influences on power output.

After reviewing the recent literature, we have discovered that there are many studies regarding power consumption but relatively few on actual range prediction. Even in that case, it is quite difficult to use those methods in the other two-wheeler e-mobility because each different e-mobility comes with a different hardware setup where technical details are mostly unknown to riders. Although some studies tried to solve it in many different ways, still more research is needed in this field that can truly address the issue.

### 3. Materials and Methods

The experiment we designed has three phases, namely, Data Collection, Data Analysis and Preparation, and Applying Machine Learning Models. Figure 1 explains our entire experimentation process. The main reason behind designing the three phases of experimentation is to work with the real data that a rider receives and/or gives. Moreover, it is imperative towards our goal, which is to provide each specific rider with a generic platform that can relay useful information with as few input data as possible without depending on the model or manufacturer of the e-mobility.



**Figure 1.** Experiment process design. See the nomenclature section for a detailed description of the variable.

An application is employed to gather raw data, which is subsequently stored in the cloud. This application may obtain information from a variety of sources, such as user interactions or through the use of APIs. Once the raw data are collected, they undergo analysis to identify patterns, followed by data preparation for the machine learning process, which involves cleaning and transforming the data. In order to evaluate the performance of various ML algorithms, 10-fold cross-validation is employed, utilizing methods such as K-nearest neighbors ( $K^*$ ), random forest (RF), support vector machine (SVM), and additive regression (AR). After applying these ML algorithms using 10-fold cross-validation, model evaluation is conducted using log performance metrics, including mean absolute ratio (MAR), root mean squared error (RMSE), and relative root squared error (RRSE). More details will be provided in the subsequent sections.

We have selected an e-scooter for the experiment called “Road Gear CT” [25], which is depicted in Figure 2. This e-scooter operates on a 36 V battery that has a capacity of 6 Ah. Additionally, it is equipped with a motor that can generate up to 300 W of power when operating at 36 V. The detailed technical description provided in Table 2 further highlights the specific capabilities of the e-scooter.



**Figure 2.** Road Gear CT electric scooter.

**Table 2.** Technical details of Road Gear CT electric scooter.

Parameter Name	Details
Battery Voltage	36 V
Battery Capacity	6 Ah
Motor Output	36 V 300 W (Rated Power)
Weight	13.5 Kg
Maximum Range	20 km
Maximum Speed	25 km/h
Wheel Size	8 inches

### 3.1. Data Collection

An application named “Ddanigo LLC. Navigation app” has been developed under the company Ddanigo LLC, 4701 Patrick Henry Dr #23, Santa Clara, CA 95054, USA [26] for collecting data and integrating machine learning techniques in the background, inspired by [8,22]. The mobile application uses flutter to ensure cross-platform availability for both Android and iOS devices.

In addition to the application, we have also compiled a comprehensive dataset containing technical information for two-wheelers worldwide. Our database contains information for 732 two-wheelers, each of which has been assigned a unique ID (named “E\_ID” in storage) for easy retrieval and analysis of technical information. This dataset encompasses a range of attributes, including the two-wheeler name, model, year, weight, voltage (V), battery capacity in watt-hours (Wh), and maximum mileage in kilometers (Km). By analyzing this diverse and extensive dataset, we aim to gain valuable insights into the performance and efficiency of various two-wheeler models.

This dataset was prepared to support the user registration process in the application. When a user registers in the application, they are required to provide details about their two-wheeler, which are then matched against the information in our database. To simplify this process for users, we only require basic information, such as the name, model, and manufacturing year, which can be easily found on the manufacturer’s website. By matching the user’s two-wheeler details with the technical information in our dataset, the application’s database creates another unique ID (named “UE\_ID” in storage) that combines both the user and two-wheeler information. This ensures that users can easily access accurate technical specifications for their two-wheeler without needing to worry about the details themselves. However, if a user’s two-wheeler is not found in the dataset, they can still register their vehicle by providing basic information. The application will then generate a unique ID, even if technical specifications are not available.

Once the two-wheeler has been set up, the application can provide navigation services. While providing navigation services, the application can collect data in the background, which are then saved to cloud storage after each individual trip. The data we have collected during each trip can be categorized into four categories: two-wheeler data, environmental data, route data, and rider behavior data. The choice to use data for range prediction is well founded, as these data categories are directly related to the factors that can influence the range of a two-wheeler. As we have discussed in Section 2, the authors [8] explored the impact of similar factors such as two-wheeler, environmental, route, and bicyclist effort data, where bicyclist effort was measured in terms of power. On the other hand, the study [22] also focused on comparable factors but they used different aspects of bicyclist performance, which is average speed. While measuring power accurately often requires additional sensors and hardware, the goal in this case is to provide a solution that does not rely on extra equipment. Instead, our focus was to utilize easily accessible information that can be collected through a smartphone. By concentrating on distance, time, and speed as rider behavior metrics, we can analyze and understand performance without the need for specialized sensors.

The experimentation was carried out only with an android device. The application requires internet connectivity and GPS to perform data collection. The routes were se-

lected based on accessibility, ability to perform rides without interruption, distinguishable differences in elevation throughout the route, and good reception of GPS and internet connectivity. Furthermore, the same route was used to obtain data for different battery levels to ensure the proper distribution of data. In Table 3, we have listed down the data that will be collected using the application from the experimental test rides, among them, which data are being collected every second of the ride and which of them are being input by the user. All the data are collected for every ride separately.

**Table 3.** List of data collected from each trip.

Category	Data	Input by User	Collected Every Second
Two-Wheeler	Initial Battery Level	Yes	No
	Remaining Battery Level	Yes	No
Environmental	Weather	No	No
	Temperature	No	No
Route	Location	No	Yes
	Elevation	No	Yes
Rider Behavior	Distance	No	Yes
	Time	No	No
	Speed	No	Yes

We collected data from approximately 100 trips conducted by a rider where the average distance and time for the trips is around 7.46 km and 24.67 min. This rider was male, around 28 years of age, and weighed around 68 kg. The data were gathered using the application, which was provided to the rider on a device for his use during various trips along the designated test routes. All the routes were covered during summer days.

The application only asks the rider to give input regarding battery levels at the start and end of each trip. The rest of the data are collected and calculated from different sensors and APIs from within the application in the background throughout the ride. We have used “Flutter Weather” plugin to collect weather type and temperature of rider’s current location when they navigate. The application was integrated with google maps API as a map server to collect route information. Furthermore, we have used direction API to obtain the list of routes. The GPS sensor along with the geocoding plugin from flutter gives us relevant information regarding location, distance, and speed every second. Subsequently, we have stored all these raw data in cloud storage for further analysis and calculation because of the restricted storage and processing capability of mobile devices [27].

The collection of personal data, such as user locations, riding habits, or other identifiable information, raises privacy concerns. To address this issue, all information related to the users was securely stored in a central database. The collected data were strictly used for the intended research purposes, and no unauthorized usage or sharing of the data was allowed. Proper access controls were implemented to ensure that only authorized personnel had access to the data for research purposes.

### 3.2. Data Analysis and Preparation

We have sorted out the dataset from the stored raw data in this phase of the experiment. Although our stored data contained data collected in a trip for every second, we have converted the data such that one trip became one instance. Firstly, we converted the battery level data we collected from the rider. We observed that the range can differ based on the initial battery state. Moreover, it was observed that, while the initial battery level was 80–100%, the range provided by e-mobility is higher, while the last 20% provides less range than other states. Based on our observation, we have categorized the initial battery levels into five categories, named A, B, C, D, and E, ranging from 80 to 100%, 60 to 79%, 40 to 59%,

20 to 39%, and 0 to 19%, respectively. Subsequently, we converted the final battery level to total battery consumption by using the following formula:

$$Battery_{con} = Battery_{init} - Battery_{final} \quad (1)$$

Although different studies have shown the impact of environmental factors coming into play for range prediction and we are collecting the environmental data for each trip, we could not use the data in our study, as the environmental data were almost constant in every single case. Therefore, we have omitted the data from our final dataset, since the impact from them will be nonexistent. Subsequently, we have also omitted the location data, as they have no factor to play in the range prediction either.

In our next step, we determined the elevation gain for the entire trip. Since we have the elevation data for every second, we needed to ascertain the total for the entire trip. To perform the calculation, we used the following Formula (2):

$$Elev_{gain} = \sum_{i=2}^n Elev_i - Elev_{i-1} [(Elev_i - Elev_{i-1}) > 0] \quad (2)$$

where  $n$  is the number of elevations in a trip.

Finally, we have converted the speed data into three fields, namely, total acceleration, total deceleration, and total stop time. In order to achieve this feat, we calculated acceleration and deceleration from the speed data of every second at first. Then, we counted the number of times a rider accelerated and decelerated in a single trip. Furthermore, the total stop time was calculated based on the number of times speed went to 0. The following formulas were used for such calculations:

$$Speed_{diff} = Speed_{curr} - Speed_{prev} \quad (3)$$

$$Acc_{total} = |A|, \text{ where } A = \{x : x \in Speed_{difference}, x > 0\} \quad (4)$$

$$Dcc_{total} = |B|, \text{ where } B = \{x : x \in Speed_{difference}, x < 0\} \quad (5)$$

$$Stop_{total} = |C|, \text{ where } C = \{x : x \in Speed_{current}, x = 0\} \quad (6)$$

Furthermore, the distance recorded was in meters, as it was collected every second from the previous position to the new position in order to clearly represent the actual travel distance rather than the distance provided by the APIs. The summation of said distance was then converted into kilometers to be included in the final dataset.

$$Dist_{total} = \frac{\sum_{i=1}^n dist_i}{1000} \quad (7)$$

where  $n$  is the number of instances in stored data.

As explained in Section 3.1, each specific two-wheeler in our database is assigned a Unique ID (named "UE\_ID" in storage) that is created by combining the specific user's information with the technical information of the specific two-wheeler. Since the user will always be making trips with their own two-wheeler, the technical details associated with the UE\_ID will remain constant. Rather than needing to include all of the technical information for the specific two-wheeler in each instance of data collected during trips, we can simply use the UE\_ID to reference the technical information stored in our database. Therefore, it is not necessary to have a lot of technical information about the two-wheelers, such as total range, battery capacity, or motor type, as the UE\_ID can be used to identify and track the vehicles accurately. Because of that, we are not including the technical details of each two-wheeler in the final version of the dataset.

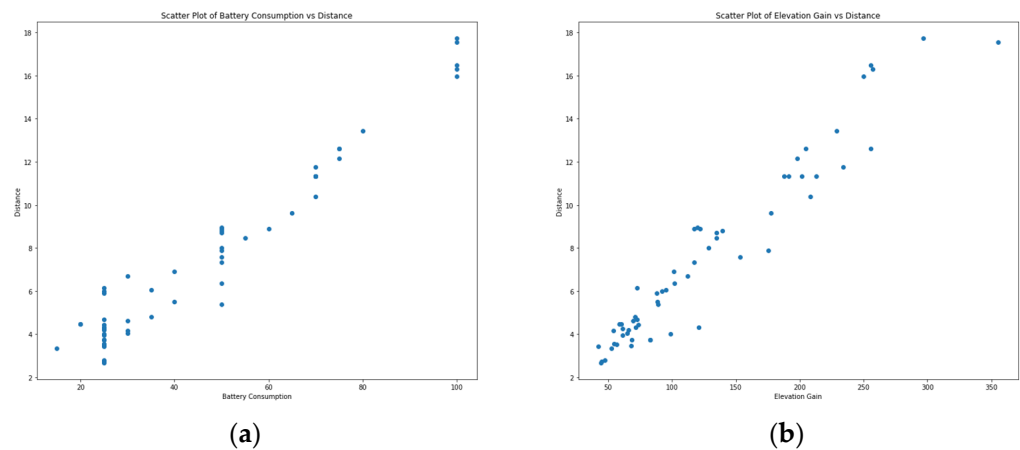
The final processed dataset, after all the processing, omission, and conversion, has 7 attributes in total, where 6 of them will be used as input features for the machine learning



models and distance will be the output. In Table 4, these attributes and their descriptions have been listed down. The dataset contains a total of 100 instances, with each instance representing a different trip. Furthermore, we have checked for irregularities and missing values in the final dataset and have found none. Figure 3a,b illustrates the scatter plot that displays the relationship between distance and the various factors of the processed dataset. Since the output variable is a continuous numerical value, regression model is a suitable choice for predicting the distancer rather than classification model. Therefore, the processed dataset was ready for the next phase.

**Table 4.** Processed dataset.

Attribute	Description
Elevation Gain	Total elevation gained in a single trip
Battery Consumption	Difference in battery level from start to end of the trip.
State of Charge	Categorization of the battery level at the start of the trip
Total Acceleration	Number of times acceleration occurred
Total Deceleration	Number of times deceleration occurred
Distance	Distance travelled in a single trip in km
UE_ID	An ID combining user and two-wheeler information

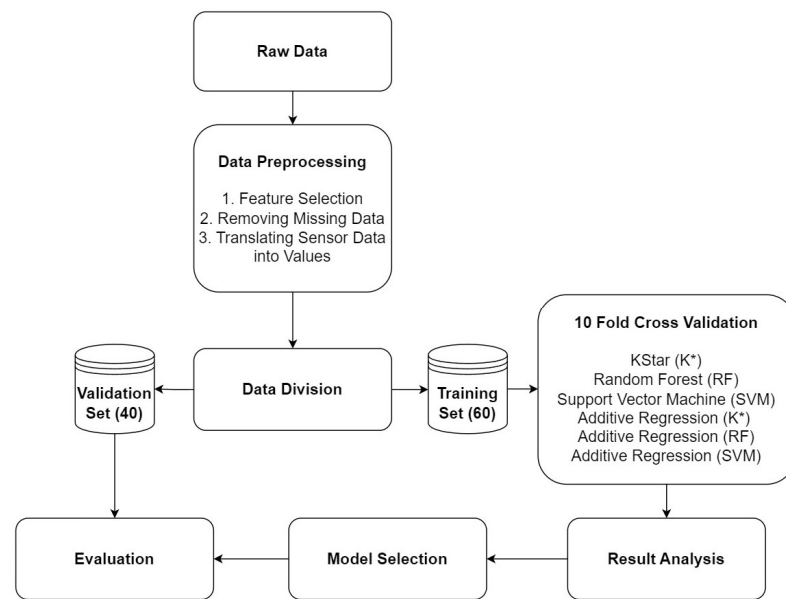


**Figure 3.** Scatter plot of input vs. output variables: (a) battery consumption vs. distance; (b) elevation gain vs. distance.

### 3.3. ML Models

Figure 4 illustrates the step-by-step process of the machine learning algorithms and their evaluation used in the study. The flowchart provides a clear visual representation of how the parameters were processed to develop the machine learning model and how it was subsequently evaluated. The process starts with the raw data, which are collected from various sources. The collected data are then preprocessed to remove any noise or unwanted information. After preprocessing, the relevant features are extracted from the data and selected based on their importance. These features are then divided into two datasets, which are training set and validation set.

The training set consists of 60 data points and is used to train the machine learning model via 10-fold cross-validation, allowing it to learn patterns and relationships between the features and the target variable. In this technique, the original dataset is randomly divided into 10 equal-sized subsets, also known as “folds”. The model is then trained on 9 of the folds and tested on the remaining fold. This process is repeated 10 times, each time using a different fold for testing and the remaining folds for training. Furthermore, the training results were used to select the best model for the validation process. The validation set, on the other hand, comprises 40 data points and is used to assess the model’s performance. The best performing model is then evaluated using the validation dataset and the evaluation matrix.



**Figure 4.** Process of the machine learning algorithm and evaluation.

In this study, we utilized six different machine learning algorithms to predict the range which is a continuous numerical value. We chose to use regression algorithms as they are commonly used to predict continuous values. In order to conduct this study, we have used java programming language with the help of the weka library. All the algorithms were coded in the same program and the evaluation was given as the output in a csv file. We have used three core algorithms and then used an additive model based on these three algorithms to make it a total of six. The algorithms we have used are KStar (K\*), random forest (RF), support vector machine (SVM), and additive regression (AR). Furthermore, we have used the K\*, RF, and SVM as the base of AR, thus rounding the total number of algorithms to six. We have used 10-fold cross-validation on these 6 algorithms and the results were stored in a csv file.

**KStar (K\*):** As a classifier, KStar relies on the similarity between training and test samples to determine the label for a given occurrence. To differentiate itself from other instance-based learners, it uses a distance function based on entropy. To assign a label to an instance, instance-based learning systems consult a database of labeled examples. The essential assumption is that analogous cases are categorized in the same way. There is a need to settle on a common understanding of “similar instance” and “similar class.” The associated parts of an instance-based learner are the distance function, which determines the degree of similarity between two instances, and the classification function, which explains how the instance similarity generates a final classification for the new instance [28].

**Random forest (RF):** To classify data, random forest uses a collection of decision trees that are generated at random. If there is use of bagging, random forests will introduce even more unpredictability. In random forests, each tree is built from a different bootstrap sample of the data, and the tree construction process is altered to improve accuracy. The optimum split across all variables is used to divide each node in a typical tree. Each node in a random forest selects a subset of predictors at random and uses the best of them to make predictions. This seemingly contradictory approach outperforms a wide variety of classifiers and support vector machines, including discriminant analysis and neural networks, and is highly resistant to overfitting [29]. RF can be described as:

$$\hat{f} = \frac{1}{B} \sum_{b=1}^B f_b(x') \quad (8)$$

**Support vector machine (SVM):** SVM-based regression models are helpful for modeling complicated relations that cannot be well captured by lower-order polynomial equations.

Because of its high power and great generalization capacity, SVM is widely used for tackling issues involving pattern recognition, classification, regression, and prediction. SVM can be described with the following equation [30]:

$$g(x) = w^T \Phi(x) + a \tag{9}$$

where  $w$  is the weight factor,  $\Phi$  represents the mapping function,  $x$  is the input vector, and  $a$  is the bias.

Additive model (AM): One type of nonparametric regression technique is the additive model (AM). The AM constructs a subset of nonparametric regression models with the help of a one-dimensional smoother [31]. Here is a simplified representation of the training procedure for additive models:

$$F_m(x) = F_{m-1}(x) + \alpha h_m(x) \tag{10}$$

Here,  $F_m(x)$  is a model ensemble that improves upon the  $F_{m-1}(x)$  model by combining  $m$  weak learning models  $h_m(x)$  to find the true solution. A scaling factor, which can range from 0 to 1, is added to reduce the contribution of each iteration in an effort to prevent the model from overfitting.

### 3.4. Model Deployment

In order to deploy our machine learning model for predicting the range of two-wheelers, we have assigned the selected model to run in the cloud. This approach allows us to take advantage of the scalability and reliability of cloud computing resources and ensures that our model is always available for use by the application’s users. Before deploying the model in the cloud, we performed extensive evaluation and testing to ensure that it meets our performance and accuracy requirements. Once the model was deemed ready for deployment, we uploaded it to our cloud infrastructure and configured it to receive user data from the application.

The ML model in the application works specifically for each user, based on their individual driving habits, road conditions, and vehicle performance history. The work process of range prediction in our application is illustrated in Figure 5. The application first recognizes the UE\_ID of the specific user who is trying to access the range prediction service. This unique ID is created by combining the user and two-wheeler information during the user registration process, which was discussed earlier in the paper. Once the UE\_ID is recognized, the application checks whether the model for that specific user’s two-wheeler is already available in the cloud or not. If the model is found, the application checks whether it needs to be updated or not. If the model needs to be updated, the application collects the latest historical data and uses it to train the model. Once the model is updated or a new model is generated, the application uses it to predict the remaining range of the user’s two-wheeler. The prediction result is then sent back to the application for display, allowing the user to plan their journey accordingly.

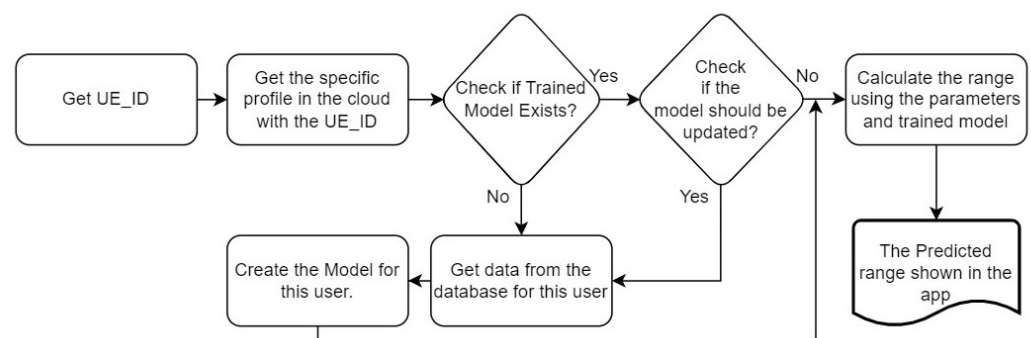


Figure 5. Flowchart of the work process of range prediction in navigation application.

#### 4. Results and Discussion

In order to determine the performances of the six machine learning algorithms, we have used mean absolute error (*MAE*), root mean squared error (*RMSE*), and root relative squared error (*RRSE*) for evaluation.

Mean absolute error (*MAE*): When comparing errors in paired observations of the same event, the mean absolute error (*MAE*) is a useful statistic. To calculate *MAE*, multiply the following formula by itself:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \tag{11}$$

Using the same measurement system as the data, we determine the mean absolute error. This accuracy statistic depends on the scale being used; hence, it cannot be used to compare series of different scales [32].

Root mean square error (*RMSE*): The root mean square error (*RMSE*) is a popular statistic for evaluating the accuracy with which a model or estimate matches up with actual data. Due to its dependence on size, *RMSE* is often employed within a single dataset to compare the accuracy of several models' forecasts rather than as a means of comparing models across datasets [33].

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \tag{12}$$

Root relative squared error (*RRSE*): A simple predictor's accuracy is quantified by its root relative squared error (*RRSE*). The mean of the observed data may be predicted with this easy-to-use metric. Total squared error is normalized by the simple predictor's total squared error, which is why we use the relative squared error. The square root of the relative squared error brings the error down to the same dimension as the expected quantity [34]. *RRSE* can be defined by the following equation:

$$RRSE = \sqrt{\frac{\sum_{i=1}^n (y' - y)^2}{\sum_{i=1}^n (y - \bar{y})^2}} \tag{13}$$

where the formula for *y* is:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y \tag{14}$$

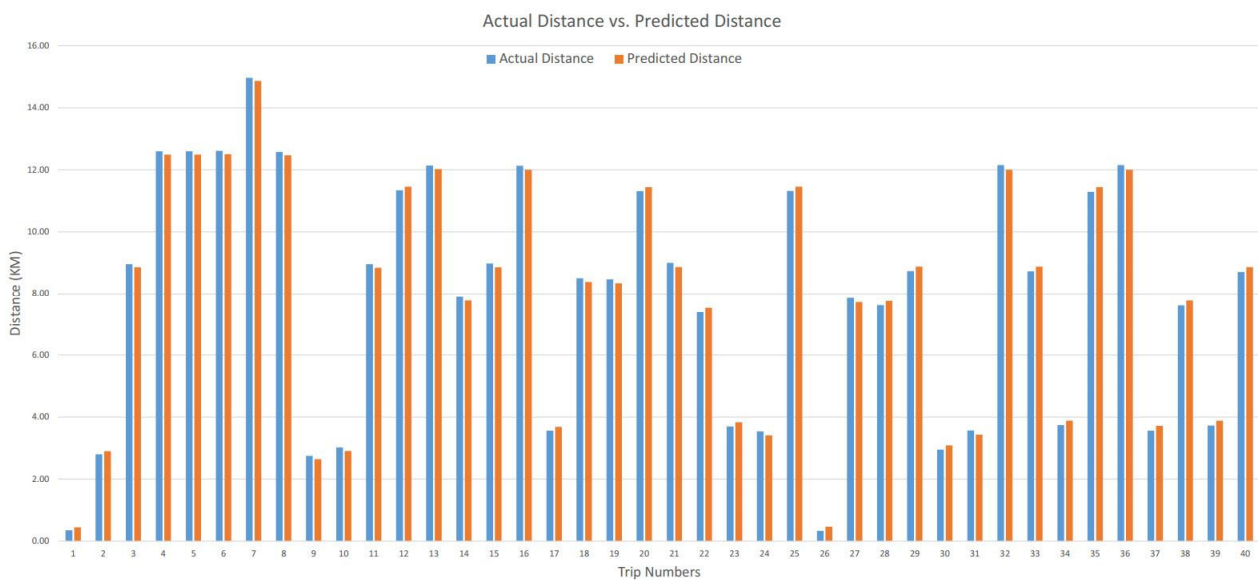
The performance metrics *MAE*, *RMSE*, and *RRSE* for the six models are compiled in Table 5. The result from Table 5 depicts that the support vector machine (SVM) algorithm produces the best results for range prediction, whereas additive models (AM) performed better than its base algorithm in all three cases. However, SVM has outperformed all others in this experiment. The high error rate is from the KStar algorithm and, given the fact that we have very limited data, that is understandable.

**Table 5.** Model performances using training dataset.

ML Model Name	MAE	RMSE	RRSE
KStar(K*)	0.26	0.4	9.65%
Random Forest (RF)	0.21	0.34	8.23%
Support Vector Machine (SMOreg)	0.15	0.29	7.03%
Additive Regression with K*	0.26	0.39	9.43%
Additive Regression with RF	0.21	0.32	7.80%
<b>Additive Regression with SMOreg</b>	<b>0.15</b>	<b>0.29</b>	<b>7.01%</b>

From Table 5, we can also see that the error rate is not that high, since this is in kilometers. We have observed from the training dataset that the mean distance we have traveled during our experiment trips is around 7.46 km, which makes our result really promising. Moreover, the best performing algorithm has an error of around 150 m on average. Therefore, it is very clear that our approach towards the problem shows much promise.

In order to evaluate the model, we have separated a dataset of 40 instances and test our best performing algorithm. Figure 6 illustrates the performance of our best model on the validation dataset by visually representing the chart that compares the actual distance travelled by the electric two-wheeler on the validation set of 40 trips with the predicted distance (in km) generated by the machine learning model. The chart shows two lines, one representing the actual distance and the other depicting the predicted distance, with trip number as the x-axis and distance (in km) travelled as the y-axis. The chart demonstrates how closely the predicted distances match the actual distances, and to identify any discrepancies or outliers that may indicate errors or limitations in the model. By comparing the two lines, we can clearly demonstrate the effectiveness of the machine learning model in predicting the distance travelled by the electric scooter.



**Figure 6.** Comparison chart of actual and predicted distance on validation set.

Figure 7 illustrates the comparison as a chart to obtain a better idea of how well our study has performed compared to others. Since our study focuses on predicting the range, we have only compared with the studies that have similar focus. The MAE is a statistical measure that indicates the average magnitude of errors in the predictions made by a model, with lower values indicating better performance. There are four studies being compared, denoted by [23,24,35], and our study. The MAE of the [23] study is 0.49, which means that the predictions made by that model had an average error of 0.49. Similarly, the MAE of the [24] study is 0.19 and the MAE of the [35] study is 0.28. Our study has an MAE of 0.13, which is the lowest value among all the studies compared. This suggests that the approach we took and the machine learning model used in our study have performed better than the models used in the other studies in terms of predicting the remaining distance of electric two-wheelers.

The Table 6 describes the parameters used in different studies for predicting the range of electric scooters. Study [23] used three different parameters, including battery and motor, scooter, and user. Study [24] used similar parameters, such as battery and motor, scooter, and user, while another study [35] only used battery parameters. In contrast, our study used three parameters, including battery, user, and road, which are state of charge, battery consumption, elevation gain, acceleration, deceleration, stop time, and

distance. By using these parameters, our study improved the accuracy of range prediction for electric two-wheelers.

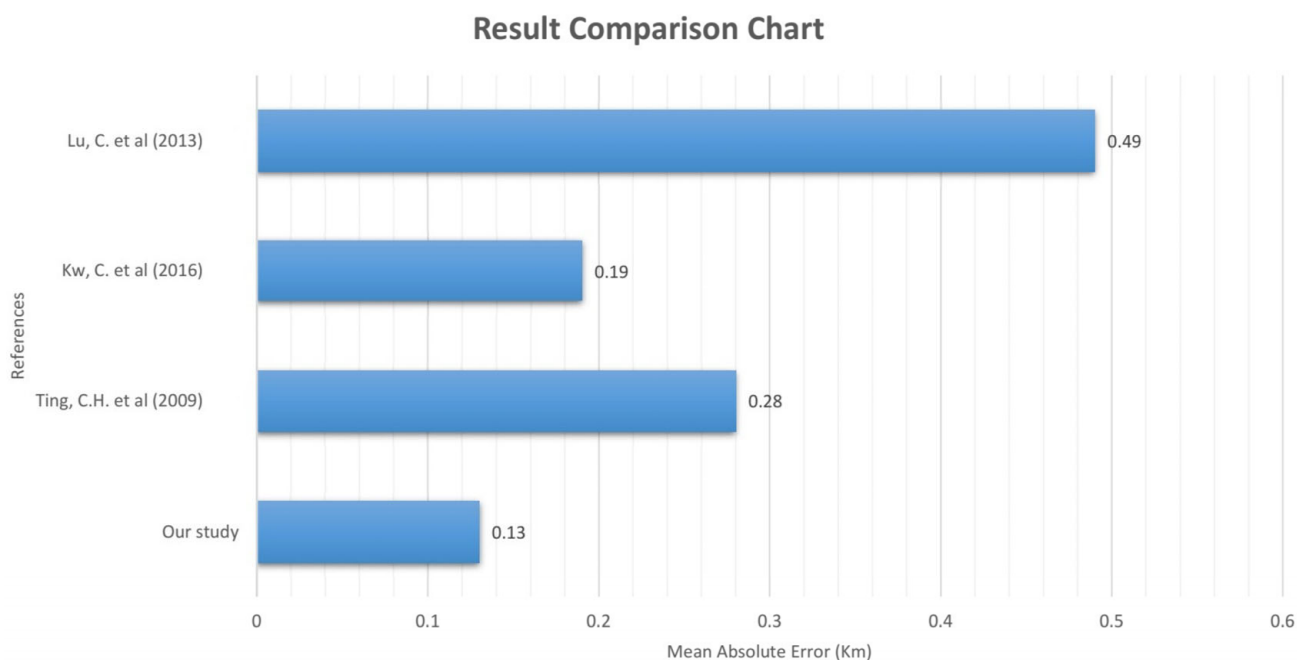
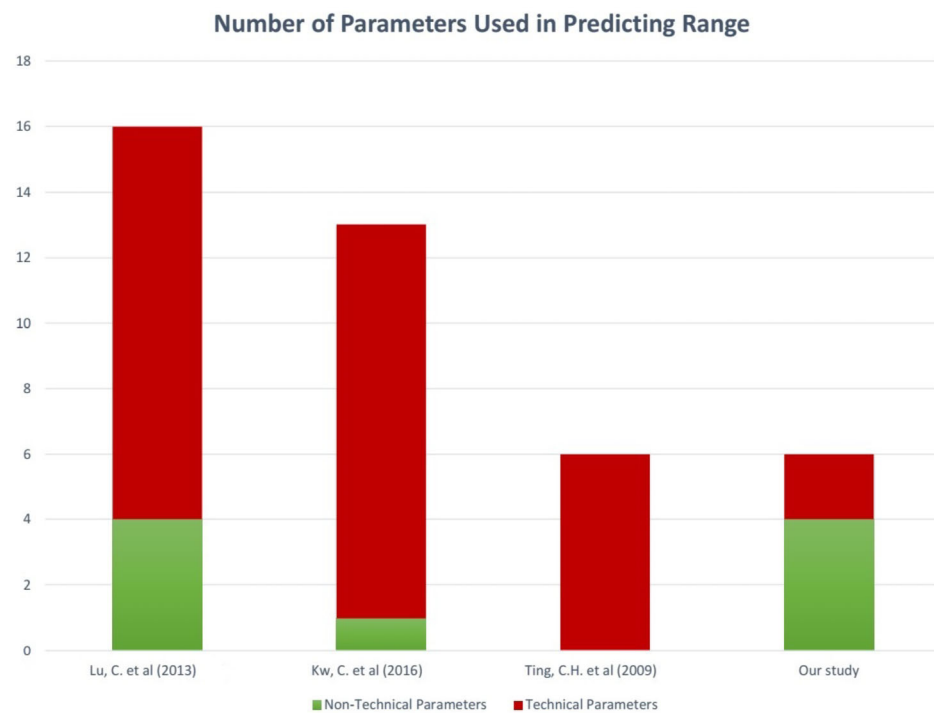


Figure 7. Evaluation with other studies Lu, C, et al. [14], Kw, C., et al. [15], Ting, C.H., et al. [35].

Table 6. Parameter lists used in other studies, including ours.

Ref.	Parameter Type	Parameter Details
[23]	Battery and Motor	Motor nominal input voltage, nominal power, nominal current, nominal battery output voltage, capacity, max continuous discharge, max charge current
	Scooter	Linear velocity, weight, rolling resistance, dynamic resistance, drag forces
	User	Weight, average speed, maximum speed, maximum acceleration
[24]	Battery and Motor	Motor power, motor torque, motor revolution speed (Basic and Max)
	Scooter	Weight, coefficient of drag, frontal area, rolling resistance, velocity, powertrain efficiency, wheel diameter, accessories load, wheel radius
	User	Speed
[35]	Battery	Discharging current, battery terminal voltage, battery capacity, time average voltage, battery temperature, scooter speed
Our study	Battery	State of charge, battery consumption (initial battery level–final battery level)
	Road	Elevation gain
	User	Acceleration, deceleration, stop time, distance

Figure 8 shows a comparison of the number of technical and nontechnical parameters used in different studies, including our one. The x-axis represents the number of studies, while the y-axis represents the number of technical and nontechnical parameters used in each study. The bars represent the total number of parameters used in each study, with green bars representing nontechnical parameters and red bars representing technical parameters. Furthermore, the parameters used in such predictions are more technical and will be very hard for the riders to provide. Nonetheless, our models have predicted quite accurately without having to rely on more technical attributes. The result shows us that we can narrow down the remaining distance to 130 m with less data and more toward user behavior. There is high potential using such a model for further studies.



**Figure 8.** Comparison of number of parameters, such as technical and nontechnical parameters, Lu, C. et al. [14], Kw, C. et al. [15], Ting, C.H., et al. [35].

## 5. Conclusions

This paper addresses the challenge of predicting the range of electric two-wheelers, which is crucial for riders who need to plan their journeys and ensure that they can reach their destination without running out of charge. The proposed solution in this paper concentrates on using machine learning techniques to predict the range. Additionally, an application has been developed for this research purpose that collects data from users and uses it to train the machine learning model. Multiple experiments have been conducted using an e-scooter for this research. We place a greater emphasis on the data that are obtained from our application using smartphone sensor or APIs and minimal engagement from the user. Therefore, the dataset we collected contains riders' behavior, road condition, weather, and battery information. Several machine learning techniques have been used for training and the results have been compared with an actual trip dataset. The method that was proposed demonstrates that it is possible to obtain an accurate range prediction even when one lacks access to specific and in-depth technical knowledge of the two-wheeler and the results are promising. The additive regression model combined with the support vector machine produced the best results of all the models. The study presented in this paper achieved a mean absolute error of 0.13 using the testing dataset, which is significantly lower than the errors reported in other studies. Moreover, the results of the predictions that were expected are consistent. Then, the selected ML model has been set on the cloud to make prediction and provide the result to our application.

We addressed the fact that our solution is able to deliver information about the range of electric two-wheelers. Because we are aware that our findings are based on a single dataset, we believe that further research involving a variety of datasets is required in order to validate the outcomes of our analysis. This is because we are aware that our conclusions are derived from a single dataset, mostly due to the fact that there was only ever one electric scooter utilized to compile that dataset. As part of our ongoing research, we intend to carry out additional studies using a variety of two-wheelers now available on the market. In addition, there is the possibility of taking into account additional parameters that have an impact on the range. However, since our ML model is user-specific, it needs to be updated regularly to ensure accurate predictions. This is because the factors that affect the range of

electric two-wheelers can vary over time and the model needs to be trained on new data to capture these changes. To address this challenge, we plan to use reinforced learning.

In conclusion, more machine learning methods are something that can be researched for this strategy. Since the fear of running out of battery power is one of the primary barriers preventing widespread use of electric vehicles, it is essential that efforts be maintained to enhance range prediction in order to make energy usage more sustainable.

**Author Contributions:** Conceptualization, A.A.; Methodology, A.A. and M.S.A.; Software, A.A. and M.S.A.; Validation, M.S.A.; Investigation, M.S.A.; Resources, A.A.; Data curation, M.S.A. and A.A.; Writing—original draft, A.A.; Writing—review & editing, M.S.A. and C.C.; Supervision, C.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data sharing not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Nomenclature

EV	Electric Vehicle
SOC	State-of-Charge
V	Voltage
Wh	Watt-hours
Km	Kilometers
$Battery_{con}$	Total Battery Consumption
$Battery_{init}$	Initial Battery Level
$Battery_{final}$	Final Battery Level
$Speed_{diff}$	Speed differences between each second
$Speed_{curr}$	Current Speed
$Speed_{prev}$	Previous Speed
$Acc_{total}$	Total Acceleration
$Dcc_{total}$	Total Deceleration
$Stop_{total}$	Total Stop Time
A	Acceleration
B	Deceleration
C	Stop Time
$Elev_{gain}$	Total Elevation Gain
$Elev_i$	Current Elevation
$Elev_{i-1}$	Previous Elevation
$Dist_{total}$	Total Distance
$dist_i$	Distance at certain point
K*	KStar
RF	Random Forest
SVM	Support Vector Machine
AM	Additive Model
AR	Additive Regression
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
RRSE	Root Relative Squared Error
E_ID	Unique ID of two-wheeler
UE_ID	Unique ID combining user and two-wheeler information



## References

- Machedon-Pisu, M.; Borza, P.N. Are Personal Electric Vehicles Sustainable? A Hybrid E-Bike Case Study. *Sustainability* **2020**, *12*, 32. [CrossRef]
- Salmeron-Manzano, E.; Manzano-Agugliaro, F. The Electric Bicycle: Worldwide Research Trends. *Energies* **2018**, *11*, 1894. [CrossRef]
- Melliger, M.A.; van Vliet, O.P.R.; Liimatainen, H. Anxiety vs Reality—Sufficiency of Battery Electric Vehicle Range in Switzerland and Finland. *Transp. Res. D Transp. Environ.* **2018**, *65*, 101–115. [CrossRef]
- Liu, H.; Li, Y.; Zhang, C.; Li, J.; Li, X.; Zhao, Y. Electric Vehicle Charging Station Location Model Considering Charging Choice Behavior and Range Anxiety. *Sustainability* **2022**, *14*, 4213. [CrossRef]
- Dill, J.; Rose, G. Electric Bikes and Transportation Policy. *Transp. Res. Rec.* **2012**, *2314*, 1–6. [CrossRef]
- Saxena, S.; Gopal, A.; Phadke, A. Electrical Consumption of Two-, Three- and Four-Wheel Light-Duty Electric Vehicles in India. *Appl. Energy* **2014**, *115*, 582–590. [CrossRef]
- MAHLE Smartbike. Available online: <https://mahle-smartbike.com/> (accessed on 18 November 2022).
- Ferreira, J.C.; Monteiro, V.; Afonso, J.A.; Afonso, J.L. Mobile Cockpit System for Enhanced Electric Bicycle Use. *IEEE Trans. Industr. Inform.* **2015**, *11*, 1017–1027. [CrossRef]
- Roberts, B.P.; Sandberg, C. The Role of Energy Storage in Development of Smart Grids. *Proc. IEEE* **2011**, *99*, 1139–1144. [CrossRef]
- De La Iglesia, D.H.; Villarubia, G.; De Paz, J.F.; Bajo, J. Multi-Sensor Information Fusion for Optimizing Electric Bicycle Routes Using a Swarm Intelligence Algorithm. *Sensors* **2017**, *17*, 2501. [CrossRef]
- Falai, A.; Giuliacci, T.A.; Misul, D.; Paolieri, G.; Anselma, P.G. Modeling and On-Road Testing of an Electric Two-Wheeler towards Range Prediction and BMS Integration. *Energies* **2022**, *15*, 2431. [CrossRef]
- Gebhard, L.; Golab, L.; Keshav, S.; Demeer, H. Range Prediction for Electric Bicycle. In Proceedings of the 7th International Conference on Future Energy Systems, e-Energy, Waterloo, ON, Canada, 21–24 June 2016; pp. 224–234. [CrossRef]
- Zhang, Y.; Wang, W.; Kobayashi, Y.; Shirai, K. Remaining Driving Range Estimation of Electric Vehicle. In Proceedings of the 2012 IEEE International Electric Vehicle Conference, IEVC, Greenville, SC, USA, 4–8 March 2012.
- Lu, C.; Kumar, P.R.; Stoleru, R.; Association for Computing Machinery. Special Interest Group on Embedded Systems; IEEE Computer Society. Technical Committee on Real-Time Systems; Association for Computing Machinery; ACM Digital Library. Real-Time Prediction of Battery Power Requirements for Electric Vehicles. In Proceedings of the 2013 ACM/IEEE International Conference on Cyber-Physical Systems (ICCPs), Philadelphia, PA, USA, 8–11 April 2013; pp. 11–20.
- Kw, C.; Yr, Y. Effectiveness Comparison of Range Estimator for Battery Electric Vehicles. *Adv. Automob. Eng.* **2016**, *5*, 839–849. [CrossRef]
- Yuan, X.; Zhang, C.; Hong, G.; Huang, X.; Li, L. Method for Evaluating the Real-World Driving Energy Consumptions of Electric Vehicles. *Energy* **2017**, *141*, 1955–1968. [CrossRef]
- Dedek, J.; Docekal, T.; Ozana, S.; Sikora, T. BEV Remaining Range Estimation Based on Modern Control Theory—Initial Study. *IIFAC-PapersOnLine* **2019**, *52*, 86–91. [CrossRef]
- Eagon, M.J.; Kindem, D.K.; Selvam, H.P.; Northrop, W.F. Neural Network-Based Electric Vehicle Range Prediction for Smart Charging Optimization. *J. Dyn. Syst. Meas. Control. Trans. ASME* **2022**, *144*, 011110. [CrossRef]
- De Cauwer, C.; Verbeke, W.; Coosemans, T.; Faid, S.; Van Mierlo, J. A Data-Driven Method for Energy Consumption Prediction and Energy-Efficient Routing of Electric Vehicles in Real-World Conditions. *Energies* **2017**, *10*, 608. [CrossRef]
- Vaz, W.; Nandi, A.K.R.; Landers, R.G.; Koylu, U.O. Electric Vehicle Range Prediction for Constant Speed Trip Using Multi-Objective Optimization. *J. Power Sources* **2015**, *275*, 435–446. [CrossRef]
- Bi, J.; Wang, Y.; Sai, Q.; Ding, C. Estimating Remaining Driving Range of Battery Electric Vehicles Based on Real-World Data: A Case Study of Beijing, China. *Energy* **2019**, *169*, 833–843. [CrossRef]
- Burani, E.; Cabri, G.; Leoncini, M. An Algorithm to Predict E-Bike Power Consumption Based on Planned Routes. *Electronics* **2022**, *11*, 1105. [CrossRef]
- Alli, G.; Formentin, S.; Savaresi, S.M. A Range-Bounding Strategy for Electric Scooters. In Proceedings of the 2012 IEEE International Electric Vehicle Conference, IEVC, Greenville, SC, USA, 4–8 March 2012.
- Yuniarto, M.N.; Wiratno, S.E.; Nugraha, Y.U.; Sidharta, I.; Nasruddin, A. Modeling, Simulation, and Validation of An Electric Scooter Energy Consumption Model: A Case Study of Indonesian Electric Scooter. *IEEE Access* **2022**, *10*, 48510–48522. [CrossRef]
- Road Gear CT. Available online: [http://www.inavi.com/Products/Sports/Gate?target=\\_roadgearCT](http://www.inavi.com/Products/Sports/Gate?target=_roadgearCT) (accessed on 26 September 2022).
- Ddanigo LLC. Navigation App. Available online: <https://www.ddanigo.com/> (accessed on 15 December 2022).
- Sadiku, M.N.O.; Musa, S.M.; Momoh, O.D. Cloud Computing: Opportunities and Challenges. *IEEE Potentials* **2014**, *33*, 34–36. [CrossRef]
- Cleary, J.G.; Trigg, L.E. K\*: An Instance-Based Learner Using an Entropic Distance Measure. In *Machine Learning Proceedings 1995, Proceedings of the Twelfth International Conference on Machine Learning, Tahoe City, CA, USA, 9–12 July 1995*; Elsevier: Atlanta, GA, USA, 1995; pp. 108–114.
- Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
- Shevade, S.K.; Keerthi, S.S.; Bhattacharyya, C.; Murthy, K.R.K. Improvements to the SMO Algorithm for SVM Regression. *IEEE Trans. Neural Netw.* **2000**, *11*, 1188–1193. [CrossRef] [PubMed]

31. Friedman, J.H. Stochastic Gradient Boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [[CrossRef](#)]
32. Willmott, C.J.; Matsuura, K. Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. *Clim. Res.* **2005**, *30*, 79–82. [[CrossRef](#)]
33. Hyndman, R.J.; Koehler, A.B. Another Look at Measures of Forecast Accuracy. *Int. J. Forecast* **2006**, *22*, 679–688. [[CrossRef](#)]
34. Root Relative Squared Error. Available online: <https://www.gepssoft.com/GeneXproTools/AnalysesAndComputations/MeasuresOfFit/RootRelativeSquaredError.html> (accessed on 10 December 2022).
35. Ting, C.H.; Tsai, C.S.; Fang, Y.L. Estimating the Residual Travel Distance of an Electrical Scooter. In Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM, Singapore, 14–17 July 2009; pp. 1348–1352.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.