Journal of Engineering Science and Technology
Vol. 20, No. 2 (2025) 490 - 509
© School of Engineering, Taylor's University

ELECTRIC VEHICLE CONSUMPTION DATASET TAILORED TO MALAYSIAN SITUATION AND IMPLEMENTED USING RAPID MINER AUTO-MODEL

NAYLI A. AZHAR[1,2], NURUL A.M. RADZI[1,2,*], VIGNA K. RAMACHANDARAMURTHY[1,2], FAIRUZ ABDULLAH[1,2], NOR H.A KAHAR[3], AZLAN AWANG[4], CHOO KAN YEEP[5]

[1]Universiti Tenaga Nasional, 43000 Kajang, Selangor, Malaysia
[2]Institute of Power Engineering (IPE), Universiti Tenaga Nasional, 43000 Kajang, Selangor, Malaysia
[3]Universiti Kuala Lumpur British Malaysian Institute, 53100, Selangor, Malaysia
[4]Universiti Teknologi Petronas, 32610, Perak, Malaysia
[5] Faculty of Engineering, Multimedia University, Persiaran Multimedia, 63100, Cyberjaya, Selangor, Malaysia
*Corresponding Author: asyikin@uniten.edu.my

Abstract

The rising global adoption of electric vehicles (EVs) is directly linked to an increase in the number of EV charging stations (EVCS). Consequently, researchers are more inclined to pursue studies in this field. However, data availability is limited, and obtaining information from relevant organizations or utilities poses significant challenges. In this study, inspired by the ACN-Data, an EV consumption dataset was created tailored to the Malaysian behaviour. Additional parameters were included to account for Malaysian behaviour, including temperature, traffic, number of EVCS available, travel distance per day (km), and EV consumption (kwh/km) based on six car brands. The dataset underwent preprocessing using Tableau Prep Builder and was then visualised using Tableau Desktop. Furthermore, the dataset was inputted into RapidMiner, which then proposed various interconnected machine learning (ML) techniques. A performance evaluation was conducted, comparing it with different ML methods. Overall, the Support Vector Machine has the best performance in term of root mean square error, absolute error, squared error and second best on relative error with 10.876, 1.278, 132.530 and 27.59%, respectively. By developing a dataset uniquely tailored to capture Malaysian behaviour for the first time, we anticipate that it will have broad and impactful applications in the future, driving significant advancements across various sectors.

Keywords: Electric vehicle, Electric vehicle consumption, Dataset, Machine learning, RapidMiner, Tableau, Visualisation.

1. Introduction

The global adoption of electric vehicles (EVs) has surged, prompting researchers to actively study and explore advancements in this rapidly evolving field. Some studies, including battery management, charging management and energy management are actively being explored in this field. This encompasses research on load forecasting for EV charging stations (EVCS), general load, and routing challenges. To address these issues effectively, a well-structured dataset is crucial for initiating comprehensive analysis.

While there are several studies that utilise datasets, there are limitations in accessing them. For instance, the dataset may be unavailable owing to privacy and confidentiality concerns, or it may be accessible online but lacking the necessary data or essential parameters. Some other issues are that for countries such as Malaysia for an example, where the implementation of EVs is still on the mid phase, the data for EVCS is limited and not publicly available.

Therefore, in order to address these issues, we have examined numerous publicly accessible datasets and decided that the ACN-Data (JPL) [1] is well-suited to serve as a source of inspiration for our proposed dataset. The ACN-Data (JPL) is made possible by a close collaboration with PowerFlex Systems which operates Adaptive Charging Networks around the United States [1]. Developing a dataset specifically designed to capture Malaysian behaviour poses significant challenges due to the scarcity of data and parameters. To address this, we propose in this paper the development of a comprehensive dataset on EV use that is uniquely tailored to Malaysian conditions. In addition, the dataset was later be used in RapidMiner, one of machine learning (ML) software to validate the performance of the dataset. The RapidMiner has been implemented in various field such as in electrical distribution substation [2], aircraft [3] and even in medical field [4].

The contributions for this paper are as follows:

- EV consumption dataset creation that are tailored to the Malaysian behaviour
- The dataset is pre-processed using Tableau Prep Builder and subsequently visualised using Tableau Desktop to enhance data visualisation.
- Using Auto Model in the RapidMiner programme, the produced dataset serves as a regression problem whereby ML predicts a qualitative, qualitative, or continuous variable depending on the chosen target variable.
- Performance evaluation was done by comparing with ML that has been suggested by the RapidMiner

The remainder of the paper is structured as follows. Section 2 provides the literature review and Section 3 discusses the dataset creation and ML process. Section 4 discusses the results obtained via the dataset creation and RapidMiner. Section 5 concludes the study.

2. Literature Review

The incorporation of EVs influences power networks by elevating demand, impacting grid stability, and need for infrastructure upgrades. Solutions like smart grids, renewable energy integration, and advanced charging networks are essential to meet these demands. This paper examines the main factors influencing EV adoption, emphasizing the roles of energy resources, and charging infrastructure. Authors examine the classifications of EV charging infrastructure, categorised by charging site and technology. A summary of the energy resources necessary for EVs is also provided and the principal characteristics of these batteries are also examined [5].

Iwafune and Kawai [6] evaluated the practical implementation of Vehicle-to-Home (V2H) systems, focussing on power conversion losses during bi-directional charging and discharging, using data from commercial Home Energy Management Systems (HEMS). It highlights the effectiveness of V2H in practical applications at a lower cost than traditional data collection methods. Results demonstrate that V2H families participated in more frequent and intensified charging activities than charging-only households. The research also developed a partial load efficiency curve, verifying that maximum efficiency was close to the rated level. Significant standby power consumption (92-142 kWh/year) was observed, and the absence of reverse power flow to the external grid from the V2H system resulted in an increase in V2H partial load operation, leading to reduced conversion efficiency [6].

Apart from that, solar-powered Distributed Generation (DG) for EV charging stations is becoming common in power systems but poses challenges like intermittent solar irradiance and peak EV loads, impacting grid stability. Amir et al., [7] proposed an Intelligent Energy Management Scheme (IEMS) with coordinated control for photovoltaic (PV)-based EV charging stations, optimizing PV and grid power usage by analysing real-time weather and demand data. The IEMS reduces peak power demand by factor two through adaptive neuro-fuzzy control for solar generation and EV load forecasting. A buffer battery system further minimizes grid impact and system losses, validated through digital simulations and real-time hardware tests.

- **EVs as Energy Storage and Perspective of Battery Swapping**

Other than that, EVs also offer eco-friendly transport and act as key energy storage systems. Using vehicle-to-grid (V2G) technology, EVs enhance grid stability by feeding energy back during peak demand. Battery swapping further boosts efficiency by enabling quick battery exchanges, reducing charging time, especially for high-demand sectors like fleets and public transport. Another study presents a hybrid energy management strategy for a Hybrid Energy Storage System (HESS) in EVs, incorporating batteries, supercapacitors (SC), and charging systems. Namib Beetle Optimisation (NBO) and Quantum Neural Networks (QNN) are integrated in the proposed NBO-QNN approach to enhance the efficiency of EV power consumption and prolong battery life [8]. The power supply and charge levels are predicted and managed by QNN, while the output voltage and current are regulated by NBO. Implemented in MATLAB, the method outperforms existing strategies like Cooperation Search Algorithm (CSA), Latent Semantic Analysis (LSA), and Grasshopper Optimization Algorithm (GOA), showing reduced stress on energy sources, better charging performance, and longer battery life, with a lower THD value [8].

Tookanlou et al. [9] presented a scheduling strategy for the charging and discharging operations of EVs, wherein each EV identifies appropriate electric vehicle charging stations (EVCSs) for these processes. The EVs organise their charging and discharging operations in accordance with the minimal driving route and economic analysis, which are determined by the prices provided by EVCSs. A bilevel optimisation problem is employed to evaluate the advantages of EVs and EVCSs in the charging and discharging of EVs. A scheduling system is proposed to facilitate communication with all EVs and EVCs, execute the bilevel optimisation problem, and transmit the results to them. The optimal electricity tariffs provided by EV charging stations for the charging and discharging of EVs are established [9].

On the other hand, Zhao et al. [10] examined distributed Aggregated battery energy storage systems (ABESSs) incorporating EV aggregators and BESSs that encounter stochastic communication problems. A distributed state-of-charge (SoC) and power balancing estimation technique utilising an event-triggered system is offered to address this issue. Unlike conventional SoC balance methods, a novel balance convergence condition addresses power imbalances between EV aggregators and BESSs. An adaptive robust estimating technique guarantees precise SoC and power evaluation amongst stochastic failures, while a distributed periodic event-triggered mechanism minimises superfluous communication and prevents the Zeno phenomenon. Simulations validate the efficacy of the technique in both charging and discharging modes [10].

- **AI Algorithms in EV Load Forecasting**

The transition to EVs requires the establishment of a charging infrastructure that is both reliable and efficient. Pardhasaradhi et al. introduces an artificial intelligence (AI)-based methodology to enhance EV infrastructure, emphasising profiling, augmentation, forecasting, explainability, and charging efficiency [11]. Profiling examines EV driver behaviours to inform strategic infrastructure development, whereas augmentation employs AI to determine ideal charging station location. Forecasting predicts future EV adoption and charging demands to aid planning. Despite recent studies, a comprehensive analysis of AI in these areas specific areas remain limited. While several authors have provided literature reviews on AI implementation, these reviews often lack a focused examination of its usage in the suggested domain. This study aims to develop and evaluate an AI framework to address this gap, validated through an empirical case study, demonstrating its effectiveness in managing dispersed energy resources for EV deployment [11].

Aduama et al. [12] proposed a forecasting technique using multi-feature data fusion to enhance the accuracy of an EV charging station load forecasting model. This approach leverages historical weather data (wind speed, temperature, humidity) as inputs to a Long Short-Term Memory (LSTM) model for robust predictions. Weather conditions significantly influence the behaviour and driving habits of EV drivers. Unlike traditional LSTM models that generate a single prediction, the proposed method makes three energy demand predictions using different multi-feature inputs, combining them through data fusion. The final model achieved a prediction error of 3.29%, outperforming standard LSTM predictions and demonstrating its potential to optimize EV charging station load forecasts.

- **Technical and Economic Analysis of EV Infrastructure**

The rise of EVs is driven by the availability of charging stations, crucial for reducing range anxiety among drivers. Sustainable charging infrastructure is necessary for the widespread proliferation of EVs since EV batteries must be recharged after a specific amount of mileage. Malaysia's National Electric Mobility Blueprint (NEMB) aims to establish itself as a centre for the EV industry by 2030, with a target of 125,000 EV charging stations. These include slow, fast, battery swap, and wireless charging stations, each with unique capabilities and challenges. This study explores these EVCS types, particularly in the context of Malaysia, examining their mechanisms, pros, cons, and related issues. Furthermore, the criteria for identifying appropriate sites for Photovoltaic Electric Vehicle Charging Stations (PEVCS) are examined, with forthcoming research employing Multi-Criteria Decision-Making (MCDM) techniques to enhance site selection [13].

The United Arab Emirates is transitioning to renewable energy because of the country's high energy consumption, fluctuating crude prices, and increasing carbon emissions. EV has the potential to enhance public health and mitigate emissions associated with climate change. The primary problem is the sustainable utilisation of renewable resources to power EVs. AlHammadi et al. analysed multiple hybrid renewable energy configurations for EV charging in Abu Dhabi, UAE, employing Hybrid Optimization of Multiple Energy Resources (HOMER) software for a techno-economic assessment [14]. The findings indicate that a model integrating solar, wind, batteries, and grid connections was ideal, generating 22,006 kWh annually at a cost of 0.06743 USD/kWh, decreasing CO_2 emissions by 384 tonnes per year, and yielding savings of 8,786.8 USD year in carbon credits. The results provide recommendations for sustainable EV charging and a robust financial rationale for investments in renewable energy [14].

- **Energy Management Strategies**

By increasing grid flexibility, plug-in hybrid electric vehicles (PHEVs) contribute to the resolution of energy and environmental issues. An innovative energy management strategy (EMS) for microgrids with renewable energy sources (RESs) and PHEVs is introduced in this study [15], which employs a self-adaptive crystal structure algorithm (SaCryStAl for short). The strategy optimizes multi-objective scheduling for microgrids with wind, solar, fuel cells, and batteries, aiming to minimize costs and environmental impacts. Three scenarios were tested: full RES integration, wind-only operation, and PHEVs in coordinated, smart, and uncoordinated charging modes. SaCryStAl outperformed other optimization methods, achieving lower costs and emissions across all scenarios, demonstrating superior efficiency compared to traditional algorithms like differential evolution and particle swarm optimization [15].

Huang et al. presents a sophisticated energy management system aimed at advancing the Green Internet of Vehicles (G-IoV) through the utilisation of intelligent edge clients and distributed EVs [16]. Hardware features, such as edge clients connected with the EV CAN bus network as an electronic control unit, and software features, like an intuitive interface, are combined in the system. Using an intelligent recommendation system, EV can optimise decisions regarding battery charging and discharging. This allows a virtual power plant (VPP) to effectively oversee the aggregated battery energy of distributed EVs, optimising the utilisation of renewable energy. Experimental findings indicate that employing federated learning to train models across EV networks surpasses direct training on individual EVs, resulting in enhanced performance [16].

In Malaysia, there exists a notable deficiency in readily accessible data concerning EV charging stations. This limitation is comprehensible, considering that Malaysia is currently in the early to mid-stages of EV adoption. The development of dataset should be guided by other accessible data or parameters, given the absence of certain crucial information. Therefore, the dataset obtained from various publicly available sources such as literature reviews, Kaggle, or Mendeley Data serve as a reference for the proposed dataset. This is critical stage on dataset creation as extraction of parameters were needed to create a suitable dataset that are relevant to Malaysian user driving behaviour. Tables 1 and 2 display a selection of publicly available datasets obtained using the specified parameter, along with the dataset proposed.

Table 1 summarizes some of literature review that utilize parameters from datasets including the proposed dataset. Our proposed dataset is inspired by JPL data due to its a real-world applicability and detail nature, which supports comprehensive analysis and modelling of EV charging behaviours.

Table 1. Literature review of studies utilizing parameters from datasets.

Dataset/Parameters	Caltech/ JPL [17]	JPL [18]	Xue et al. [19]	Feng et al. [20]	UCI [12]	Proposed
Seasonal Variation	\					\
Weather Condition	\	\			\	
Calendar Features/ Days Type	\	\				\
Arrival And Departure	\					\
Charging Session/ Charging Quantity	\					\
Policy	\					
Idle Occupancy	\					
Historical Data	\					
Temperature	\			\	\	\
Traffic Condition				\		\
Start Charging Type						
Charging Load Data		\	\	\		
Electricity Price			\			
Location	\					
Duration	\					
Charging Session Found	\					\
Number Of Sessions	\					\
Number Of Claimed Sessions	\					
Number Of Unclaimed Sessions	\					
Number Of EV Supply Equipment (EVSE)	\					\
Average Energy Consumption Per Session (Kw)	\					
Average Energy Consumption Per Day Per EVSE (Kw)	\					
Number Of Sessions Per Day	\					
Minimum Power Delivered in a Session (Kw)	\					
Maximum Power Delivered in a Session (Kw)	\					
Idle Occupied Time Per Session (Hour)	\					
Charging Start Time		\				

Charging End Time	\		
Charging Power		\	
Charging Time		\	
Power Consumption	\		
Session ID	\		
Number Of Charging Cars		\	
Speed Limit			\
Road Grade			\
Battery State-Of-Charge			\

Table 2. Some of the EV charging dataset that are publicly available including the proposed dataset.

Title, Country	Description
ACN-Data: Analysis and Applications of an Open EV Charging Dataset, United States [1]	ACN-Data was collected from two adaptive charging networks located in California which include parameters such as: connection Time, doneChargingTime, disconnectTime, kWhDelivered, siteID, stationID, sessionID, chargingCurrent, userID, kWhRequested, milesRequested, requestedDeparture
Electric Vehicle Charging Transactions, United Kingdom [21]	Charging Point Number, Transaction start, Transaction End, Charge Start, Charge End, Connect Time, Start Wh, End Wh, total Wh For City EV Data: Charging Event, Chargingport ID, Borough Operator, Plug in date and time, Unplug date and Time, Charge Start Date and Time, Charge End Date and Time, Total kWh
Electric Vehicle Charging Stations, United States [22]	Existing public EV charging stations within Connecticut from United State Department of Energy map :Station Name, Street Address, City, Access Days Time, EV Level1 EVSE Num, EV Level2 EVSE Num, EV DC Fast Count, EV Other Info and New Georeferenced Column
Electrical Vehicle Datasets, India [23]	The datasets provide valuable information on various aspects of the automobile industry and electricity consumption in the country. BehaviourSegment: Age, Profession, Marrital Status, Education, No of Dependents, Personal loan, Total Salary, and Price ChargingStationData: State/UTs and No of Charging Station EVCarsCleaned : ModelName,Brand, Price , Power(kW) and Mileage(km) EVChargingStationList: no, region, address, aux addres , latitude, longitude, type , power, and service
Electric vehicle charging station charging load data set, China [24]	Contains charging load data of three charging stations and the data provided are standardized charging load data Charging Station A/B/C: Date, Time and charging load
Proposed Dataset Inspired by [1] including taking account several parameter such as Date_Time, Day, Temperature, Number of cars,Number of Session,Travel Distance per Day (km), Traffic, Number of EVCS Available, EV consumption Increment based on temperature due to HVAC (%)EV consumption (kwh/km) (based on 6 car brand), EV consumption (kwh/km) x number of cars =[A],([A] x HVAC increment) + [A] (kwh/km) = [B], Total overall EV consumption _ [B] x travel distance (kwh)	

3. Methodology

3.1. Dataset information

In the initial stage, an in-depth analysis of EV owners' charging habits energy usage patterns will be conducted. This analysis will encompass various aspects, such as the preference for home, public, or workplace charging, as well as the specific charging times. Furthermore, the study will explore the diverse driving habits of EV owners. Additionally, the impact of these charging habits and energy usage patterns on the power grid will be thoroughly examined. Based on Fig. 1, a thorough search was conducted on multiple datasets relating to the topic. Any datasets that are not publicly available were excluded. The analysis focused on the parameter being utilised in the publicly available data set.

Since there are various publicly available EV charging datasets, certain data or parameters may not be suitable for inclusion based on Malaysian driving behaviour. Therefore, it is crucial to have essential data or parameters tailored to align with current Malaysian driving habits at this stage. Various public dataset was compared as shown in the Table 1 and Table 2. Based on these tables, the dataset was filtered out depending on the data itself (either the data is sufficient) and the parameter that related. From there, our data were inspired by the ACN- Data [1].

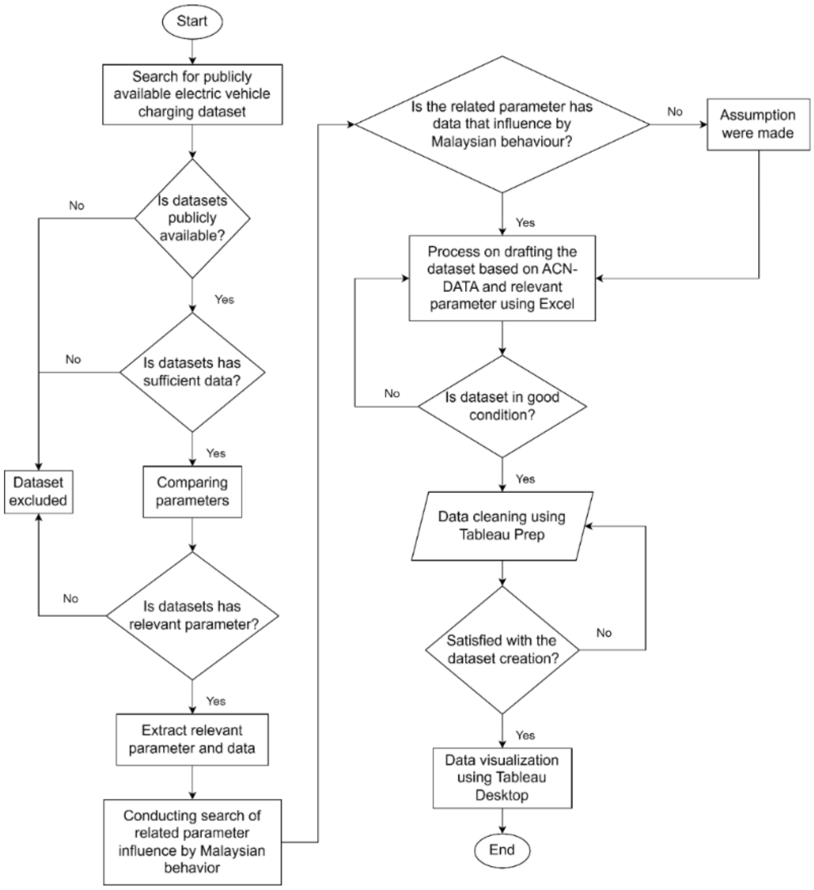

Start
Search for publicly available electric vehicle charging dataset
Is datasets publicly available?
No
Yes
Is datasets has sufficient data?
No
Yes
Dataset excluded
Comparing parameters
Is datasets has relevant parameter?
No
Yes
Extract relevant parameter and data
Conducting search of related parameter influence by Malaysian behavior
Is the related parameter has data that influence by Malaysian behaviour?
No
Assumption were made
Yes
Process on drafting the dataset based on ACN-DATA and relevant parameter using Excel
Is dataset in good condition?
No
Yes
Data cleaning using Tableau Prep
Satisfied with the dataset creation?
No
Yes
Data visualization using Tableau Desktop
End

Fig. 1. Overall data creation process.

Several assumptions were made to accommodate the behaviour of Malaysians, including that there is a rise in the number of cars and travel distance during weekends and public holidays. The public holidays considered were those observed in Malaysia in 2020. Another assumption is that the EV consumption data was based on six car brands, which are Ora Good Cat 400 Pro, BYD Atto 3 Standard, Tesla Model 3 RWD, Tesla Model Y RWD, Hyundai Ioniq 5 and BMW Ix3. These brands are chosen because due to the timeframe being set in 2020, the availability of EV car brands is significantly lower compared to the present year, given this is a limitation for the dataset creation. Therefore, we utilize six cars brands that are built according to the specifications of the year 2023. This aspect is equally important as we aim to incorporate a distinct Malaysian influence into the car brand that is specifically designed for the Malaysian market. The information below shows the dataset information and the description used in this study. Table 3 shows that the parameters alongside with the description and references that been utilised in creating the dataset specifically designed for Malaysian behaviour.

Using the data and parameters tailored to Malaysian behaviour, we created the dataset inspired by ACN-Data through the EXCEL platform. The dataset was further preprocess using Tableau Prep Builder [28]. Tableau Prep Builder is one of the data preparation tools specifically developed to assist users in cleaning, structuring, and merging data for analysis. For visualization, the cleaned dataset was fed into the Tableau Desktop. Tableau is used to allow the audience or stakeholder to have better understanding on the dataset [29].

Table 3. The parameter, description and reference of the dataset created.

Parameters	Description	Reference
Date_Time	The timestamp starts from 1/1/2020 12:00:00 AM until 31/12/2020 11:00:00 PM.	ACN-Data [1]
Day	The week begins on Monday and ends on Sunday.	-
Temperature	The temperatures are derived from historical temperature data in Malaysia in 2020	[25]
EV consumption Increment based on temperature due to HVAC (%)	Influence of temperature on energy consumption. There is an increase energy consumption (%) based on the temperature	[26]
Number of cars	Utilised the ACN-Data [1], as a baseline, we adjusted the hourly automobile count to reflect typical Malaysian habits. Assumptions were also made, such as the expectation of a rise in the number of cars on weekends and public holidays.	ACN-Data [1]
Number of Session	The number of sessions is equivalent to the number of cars. This is also utilised based on the ACN-Data	ACN-Data [1]
Travel Distance per Day (km)	The travel distance is based on assumption were in weekday, normally people travel around 20-35km daily. For weekend and public holiday there will be increases of distance travel.	-
Traffic	The traffic was categorised as low, mid, and high traffic. For example, in Malaysia, it is common for people to start work around 7 am or later, and which coincides with peak traffic hours during the weekdays.	-
Number of EVCS Available	Currently the assumption of Number of EVCS available is around 85 EVCS.	-

EV consumption (kwh/km) [based on 6 cars brands]	The initially EV consumption was calculated based on the six car brand that available in Malaysia [27]. The initial estimated parameter is essential for the subsequent step. Given the 2020 timeframe, the availability of EV car brands is significantly lower than today, limiting our dataset. Thus, we use six car brands that meet 2023 specifications.	[27]
EV consumption (kwh/km) x number of cars =[A]	This formula is to find [A] that will be used for the next formula, by taking the EV consumption (based on 6 car brand) multiply by the number of cars.	-
([A] x HVAC increment) + [A] (kwh/km) = [B]	The obtained formula [A] is later multiple with the HVAC increment and add with [A] to find [B] which shall be used to find the total overall EV consumption.	-
Total overall EV consumption _ [B] x travel distance (kwh)	The final parameter that is crucial for this study which is the total overall EV consumption that has been calculated based on previous parameter.	-

3.2. Machine learning

From the created dataset, the next process is that the dataset was fed into the Auto Model in RapidMiner. Initially the dataset needs to be imported and saved in ‘Repository.’ Repositories serve as the conventional method for storing data and other entities within RapidMiner. Data can be stored in various ways, including using the cloud, a remote RapidMiner AI Hub repository, or a local RapidMiner Studio repository, either individually or in combination. Due to its lack of integration with the file system, to retrieve data from file system or other external data sources, data must initially be imported into a repository. Fig. **2** shows the overall process in the machine leaning via the RapidMiner.

Next, we select the target, with options including prediction, clustering, and outlier detection. For this study, prediction was chosen. As the target values are numerical, RapidMiner identified this as a regression problem. Regression is another form of supervised learning. Contrary to classification, where the result is a class, regression produces a numerical value.

Based on Fig. 2, once the target for the dataset was prepared, the next step was to select the inputs. As illustrated in Fig. 3, the 'Select Input' process features color-coded status bubbles for attribute values: green indicates the highest quality, followed by yellow, and red represents the poorest quality. In this instance, RapidMiner displayed green and yellow bubbles, indicating satisfactory quality, so no changes were necessary.

Next on Select Model Types, RapidMiner recommended several MLs that are related to the problem. Since this is a regression problem, following models are recommended:

- **Generalized Linear Model (GLM):** Generalization of linear regression models. GLM offers flexibility by handling various data distributions like binomial and Poisson, making them suitable for diverse regression tasks. It provides interpretable coefficients and require no normality assumption, making them efficient and reliable for real-world applications.

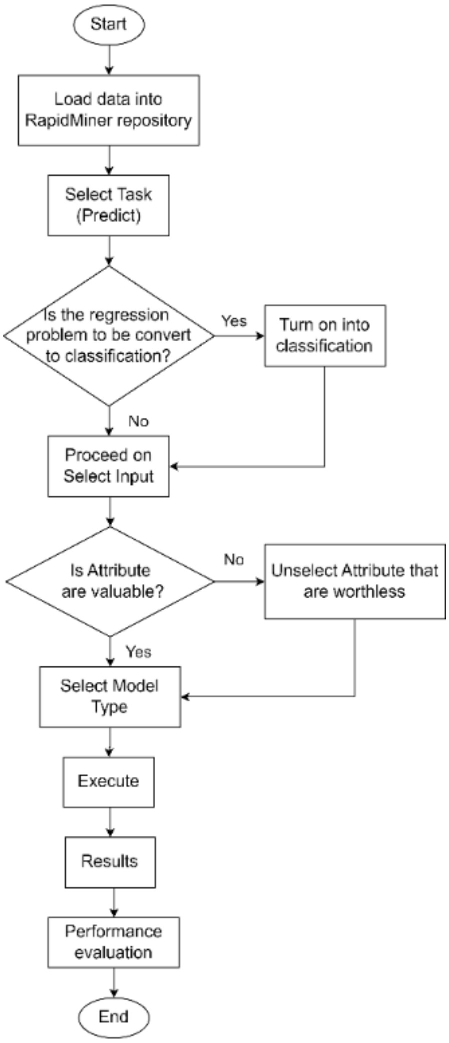
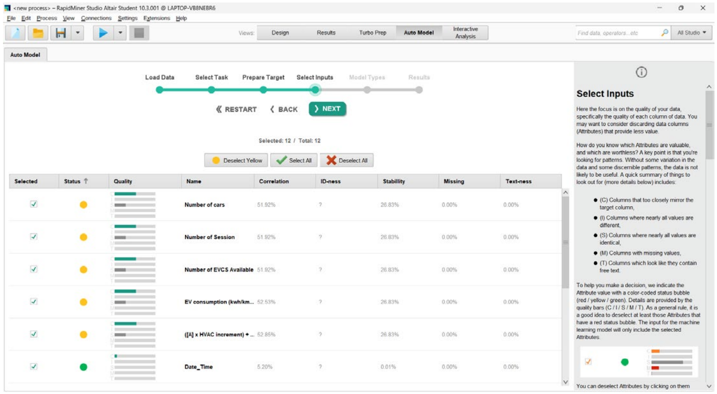

Start
Load data into RapidMiner repository
Select Task (Predict)
Is the regression problem to be convert to classification?
Yes
Turn on into classification
No
Proceed on Select Input
Is Attribute are valuable?
No
Unselect Attribute that are worthless
Yes
Select Model Type
Execute
Results
Performance evaluation
End

Fig. 2. Machine learning process via RapidMiner.


<new process> – RapidMiner Studio Altair Student 10.3.001 @ LAPTOP-VB8NEBR6
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Interactive Analysis
Find data, operators...etc
All Studio
Auto Model
Load Data
Select Task
Prepare Target
Select Inputs
Model Types
Results
RESTART
BACK
NEXT
Deselect Yellow
Select All
Deselect All

Selected: 12 / Total: 12

Selected	Status	Quality	Name	Correlation	ID-ness	Stabillity	Missing	Text-ness
☑	●		Number of cars	51.92%	?	26.83%	0.00%	0.00%
☑	●		Number of Session	51.92%	?	26.83%	0.00%	0.00%
☑	●		Number of EVCS Available	51.92%	?	26.83%	0.00%	0.00%
☑	●		EV consumption (kwh/km...	52.53%	?	26.83%	0.00%	0.00%
☑	●		([A] x HVAC increment) + ...	52.85%	?	26.83%	0.00%	0.00%
☑	●		Date_Time	5.20%	?	0.01%	0.00%	0.00%

Select Inputs

Here the focus is on the quality of your data, specifically the quality of each column of data. You may want to consider discarding data columns (Attributes) that provide less value.

How do you know which Attributes are valuable, and which are worthless? A key point is that you're looking for patterns. Without some variation in the data and some discernible patterns, the data is not likely to be useful. A quick summary of things to look out for (more details below) includes:

- (C) Columns that too closely mirror the target column,
- (I) Columns where nearly all values are different,
- (S) Columns where nearly all values are identical,
- (M) Columns with missing values,
- (T) Columns which look like they contain free text.

To help you make a decision, we indicate the Attribute value with a color-coded status bubble (red / yellow / green). Details are provided by the quality bars (C / I / S / M / T). As a general rule, it is a good idea to deselect at least those Attributes that have a red status bubble. The input for the machine learning model will only include the selected Attributes.

You can deselect Attributes by clicking on them

Fig. 3. The select input in auto model RapidMiner.

- **Deep Learning (DL):** Multilayered neural network-based machine learning. It can generate novel features from limited training data, combined with continuous training, making its architecture adaptable to changes and capable of addressing a variety of challenges. It also produces dependable and actionable results for tasks using unsupervised learning approaches [30].
- **Decision Tree (DT):** Flowchart-like structure used for decision-making and data analysis. The implementation produces favourable outcomes even with limited data, and it is straightforward to maintain and user-friendly, necessitating only a minimal number of parameters [31].
- **Random Forest (RF):** Combines the output of multiple DTs to reach a single result. It is resistant to overfitting, achieves high accuracy, and performs well with many variables and large datasets. Additionally, it automatically handles missing values and does not require variable transformation [31].
- **Gradient Boosted Trees (GBT):** A robust yet intricate model employing ensembles of DTs. It is more accurate compared to other models and train faster, particularly on larger datasets. It offers support for handling categorical features, and some can handle missing values natively.
- **Support Vector Machine (SVM):** It is versatile and may be utilised for both classification and regression problems across various applications. It is memory efficient in high-dimensional spaces when the number of features (N) exceeds the number of samples. It performs exceptionally well with clearly separated margins and automatically generates non-linear functions [31].

After the execution, the performance was evaluated and compared between the ML models. This performance evaluation will be discussed in detail in the next section.

4. Results and Discussion

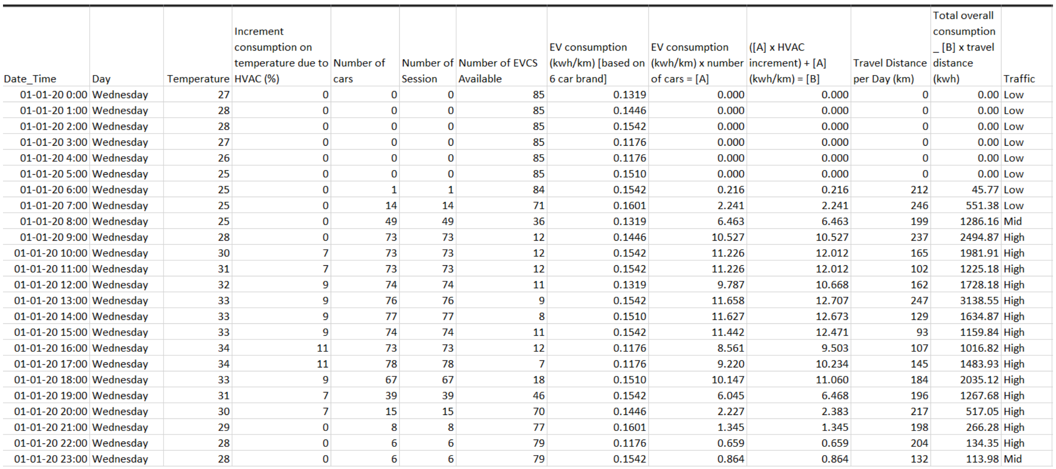
4.1. Dataset information

The initial step is to establish a dataset based on the gathered data on user behaviour from public available data. From the gathered parameters, an hourly dataset was created by Excel and contained several columns including the Date Time, number of cars, temperature, travel distance, traffic, total EV consumption and others related, as shown in Table 4. The time span of the dataset is from January 2020 until December 2020. The dataset is based on the demographics of Selangor, Malaysia, which is a realistic urban area with a high density of EVCS and EV. It provides a solid foundation for future comparisons with real-world data in our research.

Table 4. Dataset information.

Dataset Information	
Date_Time [1]	Traffic
Day	Number of EVCS Available
Temperature [25]	EV consumption (kwh/km) [based on 6 cars brands]
EV consumption Increment based on temperature due to HVAC (%) [26]	EV consumption (kwh/km) x number of cars =[A] [27]
Number of cars [1]	([A] x HVAC increment) + [A] (kwh/km) = [B]
Number of Session [1]	Total overall EV consumption _ [B] x travel distance (kwh)
Travel Distance per Day (km)	

Fig. **4** shows a snapshot of the created dataset. Dataset consists of 13 columns and 8785 rows. The parameters were finalized from various publicly available dataset and this dataset is inspired from the ACN-Data (JPL) [1]. Several assumptions were made to accommodate the behaviour of Malaysians, including that there is a rise in the number of cars and travel distance during weekends and public holidays. The public holidays considered were those observed in Malaysia in 2020. In overall, the dataset consists of 13 columns and 8785 rows, capturing hourly data from 1st January 2020 to 31st December 2020. It is specifically based on the demographics of Selangor, Malaysia, allowing for targeted analysis and modelling.



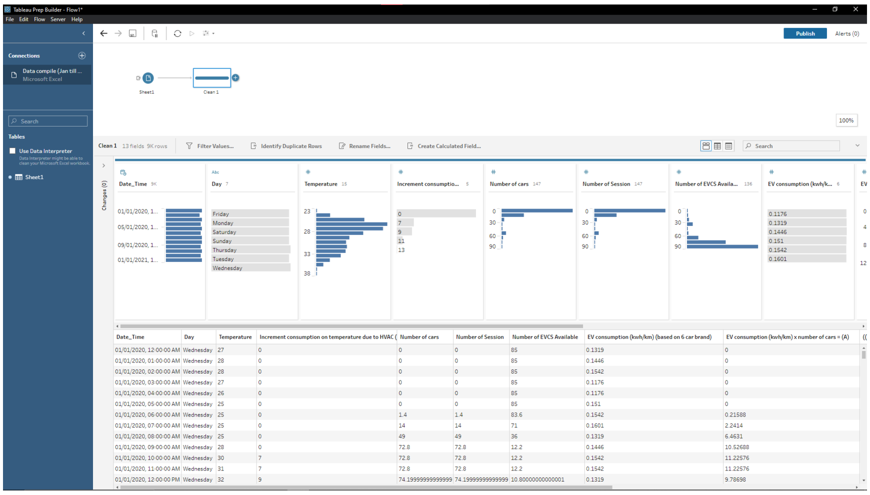
Date_Time	Day	Temperature	Increment consumption on temperature due to HVAC (%)	Number of cars	Number of Session	Number of EVCS Available	EV consumption (kwh/km) [based on 6 car brand]	EV consumption (kwh/km) x number of cars = [A]	([A] x HVAC increment) + [A] (kwh/km) = [B]	Travel Distance per Day (km)	Total overall consumption _ [B] x travel distance (kwh)	Traffic
01-01-20 0:00	Wednesday	27	0	0	0	85	0.1319	0.000	0.000	0	0.00	Low
01-01-20 1:00	Wednesday	28	0	0	0	85	0.1446	0.000	0.000	0	0.00	Low
01-01-20 2:00	Wednesday	28	0	0	0	85	0.1542	0.000	0.000	0	0.00	Low
01-01-20 3:00	Wednesday	27	0	0	0	85	0.1176	0.000	0.000	0	0.00	Low
01-01-20 4:00	Wednesday	26	0	0	0	85	0.1176	0.000	0.000	0	0.00	Low
01-01-20 5:00	Wednesday	25	0	0	0	85	0.1510	0.000	0.000	0	0.00	Low
01-01-20 6:00	Wednesday	25	0	1	1	84	0.1542	0.216	0.216	212	45.77	Low
01-01-20 7:00	Wednesday	25	0	14	14	71	0.1601	2.241	2.241	246	551.38	Low
01-01-20 8:00	Wednesday	25	0	49	49	36	0.1319	6.463	6.463	199	1286.16	Mid
01-01-20 9:00	Wednesday	28	0	73	73	12	0.1446	10.527	10.527	237	2494.87	High
01-01-20 10:00	Wednesday	30	7	73	73	12	0.1542	11.226	12.012	165	1981.91	High
01-01-20 11:00	Wednesday	31	7	73	73	12	0.1542	11.226	12.012	102	1225.18	High
01-01-20 12:00	Wednesday	32	9	74	74	11	0.1319	9.787	10.668	162	1728.18	High
01-01-20 13:00	Wednesday	33	9	76	76	9	0.1542	11.658	12.707	247	3138.55	High
01-01-20 14:00	Wednesday	33	9	77	77	8	0.1510	11.627	12.673	129	1634.87	High
01-01-20 15:00	Wednesday	33	9	74	74	11	0.1542	11.442	12.471	93	1159.84	High
01-01-20 16:00	Wednesday	34	11	73	73	12	0.1176	8.561	9.503	107	1016.82	High
01-01-20 17:00	Wednesday	34	11	78	78	7	0.1176	9.220	10.234	145	1483.93	High
01-01-20 18:00	Wednesday	33	9	67	67	18	0.1510	10.147	11.060	184	2035.12	High
01-01-20 19:00	Wednesday	31	7	39	39	46	0.1542	6.045	6.468	196	1267.68	High
01-01-20 20:00	Wednesday	30	7	15	15	70	0.1446	2.227	2.383	217	517.05	High
01-01-20 21:00	Wednesday	29	0	8	8	77	0.1601	1.345	1.345	198	266.28	High
01-01-20 22:00	Wednesday	28	0	6	6	79	0.1176	0.659	0.659	204	134.35	High
01-01-20 23:00	Wednesday	28	0	6	6	79	0.1542	0.864	0.864	132	113.98	Mid

Fig. 4. A snapshot of dataset creation based on Malaysian behaviour.

Although the created dataset was inspired from the ACN-Data (JPL) [1], several assumptions were made to reflect Malaysian behaviour; introducing certain validity risks. Additionally, the dataset's regional and temporal scope could limit the generalizability of findings. Although multiple models were employed to enhance the study, the potential for overfitting or underfitting remains, which could impact the interpretation of performance indicators. However, we have taken steps to mitigate these risks by implementing robust model validation techniques, such as cross-validation and regularization, to ensure reliable results. In addition, we prioritized proper management of the proposed dataset, following best practices in data handling to maintain accuracy and uphold ethical research standards. Acknowledging these limitations is crucial for understanding the study's findings, and we are committed to addressing them in future research to further refine and improve the analysis.

4.2. Tableau prep builder

For dataset curation, the dataset was further prepared using the Tableau Prep Builder [28]. With Tableau Prep Builder, a more contemporary method of data preparation is provided, which speeds up the process of combining, shaping, and cleaning data in preparation for analysis. By providing a transparent and direct approach to data prep, quality data can be received in a few clicks. This step-in data curation is critical as to ensure the quality, usefulness, compliance, and efficiency of data. Fig. **5** shows a snapshot of Tableau Prep Builder.



Date_Time	Day	Temperature	Increment consumption on temperature due to HVAC (	Number of cars	Number of Session	Number of EVCS Available	EV consumption (kwh/km) (based on 6 car brand)	EV consumption (kwh/km) x number of cars = (A)	((
01/01/2020, 12:00:00 AM	Wednesday	27	0	0	0	85	0.1319	0	
01/01/2020, 01:00:00 AM	Wednesday	28	0	0	0	85	0.1446	0	
01/01/2020, 02:00:00 AM	Wednesday	28	0	0	0	85	0.1542	0	
01/01/2020, 03:00:00 AM	Wednesday	27	0	0	0	85	0.1176	0	
01/01/2020, 04:00:00 AM	Wednesday	26	0	0	0	85	0.1176	0	
01/01/2020, 05:00:00 AM	Wednesday	25	0	0	0	85	0.151	0	
01/01/2020, 06:00:00 AM	Wednesday	25	0	1.4	1.4	83.6	0.1542	0.21588	
01/01/2020, 07:00:00 AM	Wednesday	25	0	14	14	71	0.1601	2.2414	
01/01/2020, 08:00:00 AM	Wednesday	25	0	49	49	36	0.1319	6.4631	
01/01/2020, 09:00:00 AM	Wednesday	28	0	72.8	72.8	12.2	0.1446	10.52688	
01/01/2020, 10:00:00 AM	Wednesday	30	7	72.8	72.8	12.2	0.1542	11.22576	
01/01/2020, 11:00:00 AM	Wednesday	31	7	72.8	72.8	12.2	0.1542	11.22576	
01/01/2020, 12:00:00 PM	Wednesday	32	9	74.19999999999999	74.19999999999999	10.80000000000001	0.1319	9.78698	

Fig. 5. A snapshot of Tableau prep builder.

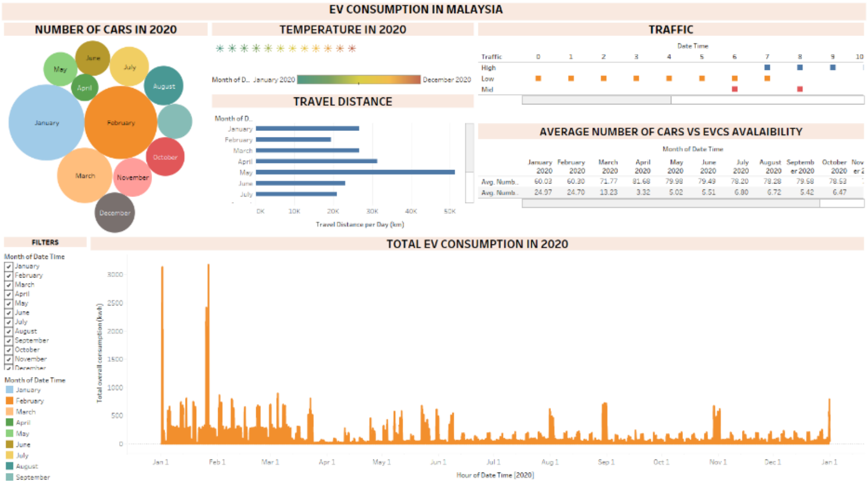
4.3. Tableau desktop

Upon data curation, for better dataset visualization, the data were fed into Tableau Desktop [29]. Tableau Desktop provides all the necessary tools to access, visualise, and analyse the data. By utilising a user-friendly drag and drop interface, one may efficiently reveal concealed insights to expedite significant business choices, even in the absence of an internet connection including in a secure safe environment. Another public and free version is called Tableau Public, where it does have some limitation compared to the Tableau Desktop. In both versions, there is a dashboard. A dashboard primarily serves the purpose of displaying a complete synopsis of data derived from various sources. The monitoring, measuring, and analysis of pertinent data in critical areas can be accomplished with the help of dashboards. Fig. **6** shows the overview of dashboard for the Malaysia EV consumption in 2020.

From Fig. 6, numerous charts and tables were included to enhance the visualisation of the created dataset. The total EV consumption graph displayed on the dashboard is generated by considering several factors such as number of cars, temperature, travel distance and others related. Each parameter is interconnected with the others. The inclusion of the filtering feature in the dashboard is essential for data analysis. Various sorts of filters exist, but in this instance, the specific one employed was date filtering. Tableau also offers "Filter Actions" to improve dashboard interactivity. These actions enable users to filter an entire dashboard based on the value selected in a specific visualization.

4.4. Machine learning via RapidMiner

This concludes the Auto Model process, which allows for the examination of the generated models in conjunction with other results where the outcome is determined by the data and decisions made. In Auto Model process, outcomes are computed in the background. Nevertheless, evaluation of the results may commence promptly upon their completion. RapidMiner does not endorse the use of black box. Hence, it is possible to consistently access the process that generated the model and its associated outcomes. Making the RapidMiner Auto Model is transparent, ensuring the results are interpretable and trustworthy.



Date Time

Traffic	0	1	2	3	4	5	6	7	8	9	10
High								■	■	■	
Low	■	■	■	■	■	■	■	■			
Mid							■		■		

AVERAGE NUMBER OF CARS VS EVCS AVALAIBILITY

Month of Date Time

	January 2020	February 2020	March 2020	April 2020	May 2020	June 2020	July 2020	August 2020	September 2020	October 2020	Nov er 2
Avg. Numb..	60.03	60.30	71.77	81.68	79.98	79.49	78.20	78.28	79.58	78.53	
Avg. Numb..	24.97	24.70	13.23	3.32	5.02	5.51	6.80	6.72	5.42	6.47	

Fig. 6. The created dashboard of EV consumption in Malaysia via Tableau desktop.

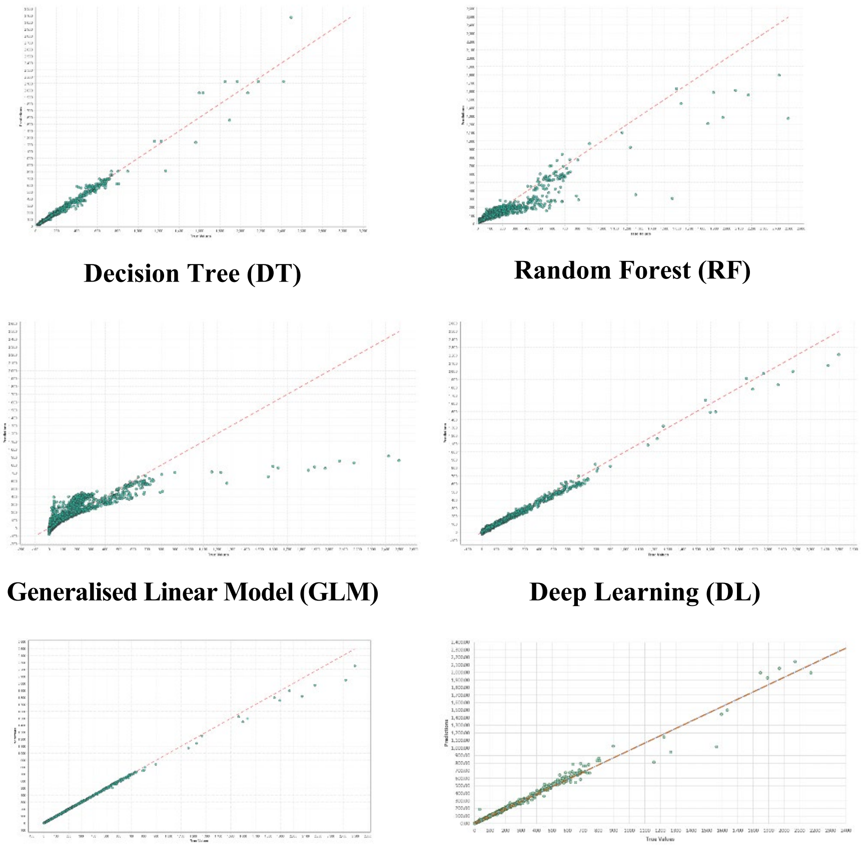
Root mean square error (RMSE), absolute error (the average absolute deviation of the prediction from the actual value), relative error (the average relative error), and square error (the averaged squared error) are the four performance metrics that can be used for comparisons. Apart from that, RapidMiner also measures overall time, training time, and scoring time. But since this study only considers one time running scenario hence the overall time, training time and scoring time is not evaluated.

Based on Table 5, it can be observed that the SVM shows the lowest RMSE, absolute error and squared error, and second lowest in terms of relative error having 10.88, 1.28, 132.53 and 27.59% respectively. The second highest model is GBT with 13.09, 2.73, 179.39 and 28.66% for RMSE, absolute error, squared error, and relative error. SVM shows the lowest average value compared to other ML models studied. Low error values indicate that the model can make highly precise predictions and effectively matches the given data. SVM has lower values due to it exhibit resilience when dealing with complex and high-dimensional issues, while also requiring a limited number of samples. In addition, the formulation incorporates the notion of minimising structural risk and is proven to be more effective than the standard principle of minimising empirical risk used in conventional machine learning methods [32].

It is stated that, in summary, Auto Model utilises 60% of the original dataset for training the models, which is commonly known as the training set [33]. The test set consists of the remaining 40% of the initial dataset and is used to evaluate the performance of the models. To facilitate understanding, a visual representation comparing predicted values to actual values was created for 40% of the validation scenarios where the actual values are known. This is illustrated in the predictions chart shown in Fig. 7. Every point on the graph corresponds to a particular predicted and its actual value. A better model is indicated by the proximity of the dots to the orange line.

Table 5. Performance comparison for GLM, DL, DT, RF, GBT and SVM.

Model	Root Mean Squared Error	Absolute Error	Squared Error	Relative Error
Generalized Linear Model (GLM)	84.34	34.29	7248.23	66.91%
Deep Learning (DL)	17.37	10.14	304.99	45.02%
Decision Tree (DT)	22.07	4.84	516.57	3.07%
Random Forest (RF)	64.34	26.80	4342.90	51.80%
Gradient Boosted Trees (GBT)	13.09	2.73	179.39	28.66%
Support Vector Machine (SVM)	10.88	1.28	132.53	27.59%
Average	**35.35**	**13.35**	**2120.77**	**37.18%**



Decision Tree (DT)

Random Forest (RF)

Generalised Linear Model (GLM)

Deep Learning (DL)

Support Vector Machine (SVM)

Gradient Boosted Tree (GBT)

Fig. 7. Comparison based on the prediction charts.

Figure 7 shows the prediction charts of DT, RF, GL, DL and SVM. The GBT prediction charts were not included as it has been a known issue, and may be fixed in a recent release (current RapidMiner version is 10.3.001) [34]. However, given prediction data by RapidMiner, the prediction graph was manually created using Excel. From the prediction graph, SVM, GBT and DL have high prediction compared to DT, RF and GLM, with SVM being the best. This can be validated with the performance evaluation done earlier where RF and GLM have the worst performance in terms of RMSE, absolute error and squared error.

When comparing the performance of the SVM with other models, it is evident that SVM outperforms its counterparts, including GBT, DL, DT, RF, and GLM. The results indicate that SVM consistently yields the lowest RMSE, absolute error, and squared error, positioning it as the most accurate model in the analysis. In contrast, RF and GLM exhibit the highest error metrics, indicating their limited effectiveness in capturing the underlying patterns of the dataset. Although GBT and DL also demonstrate respectable performance, they do not match the precision of SVM. This is due to SVMs handle non-linear relationships well through the use of kernel functions, fitting complex patterns without the high variance associated with more complex models like DL. They typically require less hyperparameter tuning, resulting in more stable performance across different datasets, whereas inconsistent tuning in DL and GBT can lead to higher error rates.

Moreover, SVM achieved a second place ranking for relative error, reflecting its robustness not only in terms of overall accuracy but also in handling varying measurement scales effectively. This thorough analysis underscores the dependability of the SVM model in assessing the dataset using Auto Model, confirming its appropriateness for performance validation. Additionally, the proficiency of SVM and other high-performing models in identifying intricate patterns within the data enhances our understanding of the dataset's strengths and limitations, which is essential for guiding future research endeavours and optimizing model performance in similar applications. This comparative analysis not only emphasizes the superiority of SVM in validating the dataset's performance but also points to the need for further exploration of its application in similar datasets and scenarios.

The validation using RapidMiner Auto Model, effectively showcases the performance of models on the proposed dataset. Overall, the EV Consumption Dataset is also highly significant for researchers and the automotive industry. For the research community, this dataset provides a valuable foundation for studying EV usage patterns, grid impact, and user behaviour in urban settings, which can advance predictive modelling, optimization algorithms, and policy design. For the automotive industry, insights from this dataset can guide the development of EV infrastructure, enhance user experience through data-driven service improvements, and help in assessing market demand and consumption trends. This broader impact underscores the dataset’s importance beyond basic validation.

5. Conclusion

In conclusion, the growing global adoption of EVs is directly related to an increase in the number of EVCS. As a result, researchers are more likely to undertake a new study in the same field. Nevertheless, the availability of data is restricted and obtaining information from the relevant organisation or utility is challenging. This is challenging especially in a country that is in the mid phase on implementing EVs such as Malaysia.

The necessity of doing research in this subject is vital for the advancements of future development. Therefore, we rely on the publicly available dataset. However, these publicly available data is in their current state form and manual extraction needs to be done to extract some of the information. Furthermore, the process of creating a dataset should be guided by other accessible data or parameters, considering the absence of some necessary information.

Drawing inspiration from the ACN-Data, this work intends to construct an EV consumption dataset specific to Malaysian behaviour. Several parameters were incorporated into the dataset, specifically customised to reflect Malaysian behaviour, including factors such as temperature, traffic, and others. The dataset was then pre-processed via Tableau Prep Builder and visualised using the Tableau desktop.

The performance validation of the dataset was done using AutoModel in the RapidMiner. RapidMiner listed several ML that are suitable for the problem for such DT, RF, GLM, DL, SVM and GBT. Since this study considered as regression problem, the performance evaluated are in term of RMSE, absolute error, squared error, and relative error. By developing a dataset specifically designed to capture Malaysian behaviour, we anticipate that this dataset will have broader applications in the future.

Acknowledgement

We would like to acknowledge the Government Linked University (GLU) grant with grant number of 052023GLU.

References

1. Lee, Z.J.; Li, T.; and Low, S.H. (2019). ACN-Data: Analysis and applications of an open EV charging dataset. *Proceedings of the Tenth ACM International Conference on Future Energy Systems*, Phoenix, Arizona, USA, 139-149.
2. Azhar, N.A.; Radzi, N.A.M.; Azmi, K.H.M.; Samidi, F.S.; and Zainal, A.M. (2022). Criteria selection using machine learning (ML) for communication technology solution of electrical distribution substations. *Applied Sciences*, 12(8), 3878.
3. Marzukhi, S.; Awang, N.; Alsagoff, S.N.; and Mohamed, H. (2021). RapidMiner and machine learning techniques for classifying aircraft data. *Journal of Physics: Conference Series,* 1997(1), 12012.
4. Surojudin, N.; Ermanto, E.; Danny, M.; and Pratama, S. (2024). Implementation of the Naive bayes algorithm for death due to heart failure using rapid miner. *Brilliance Research of Artificial Intelligence*, 4(1), 294-302.
5. Aduama, P.; Al-Sumaiti, A.S.; and Al-Hosani, K.H. (2023). Electric vehicle charging infrastructure and energy resources: A review. *Energies*, 16(4), 1965.
6. Iwafune, Y.; and Kawai, T. (2024). Data analysis and estimation of the conversion efficiency of bidirectional EV chargers using home energy management systems data. *Smart Energy*, 15, 100145.
7. Amir, M. et al. (2024). Intelligent energy management scheme-based coordinated control for reducing peak load in grid-connected photovoltaic-powered electric vehicle charging stations. *IET Generation, Transmission & Distribution*, 18(6), 1205-1222.

8. AlKawak, O.A.; Kumar, J.R.R.; Daniel, S.S.; and Reddy, C.V.K. (2024). Hybrid method-based energy management of electric vehicles using battery-super capacitor energy storage. *Journal of Energy Storage*, 77, 109835.
9. Tookanlou, M.B.; Marzband, M.; Al-Sumaiti, A.S.; and Asadi, S. (2021). *Energy vehicles as means of energy storage: impacts on energy markets and infrastructure*. In Mohammadi-Ivatloo, B.; Shotorbani, M.A.; and Anvari-Moghaddam, A. (Eds.), *Energy Storage in Energy Markets.* Academic Press, 131-146.
10. Zhao, J.-W.; Zhang, H.-L.; and Wang, C. (2024). Distributed state-of-charge and power balance estimation for aggregated battery energy storage systems with EV aggregators. *Energy*, 305, 132193.
11. Pardhasaradhi, B. et al. (2024). Intelligent integration: Harnessing artificial intelligence for enhances performance and efficiency in electric vehicles. *Journal of Electrical Systems*, 20(5s), 376-385.
12. Aduama, P.; Zhang, Z.; and Al-Sumaiti, A.S. (2023). Multi-feature data fusion-based load forecasting of electric vehicle charging stations using a deep learning model. *Energies*, 16(3), 1309.
13. Syahirah N.A.; and Farah R.N. (2024). Charging ahead: Statistics on electric vehicle charging station allocation and uptake trends in Malaysia. *Applied Mathematics and Computational Intelligence* (*AMCI*), 13(1), 69-083.
14. AlHammadi, A. et al. (2022). Techno-economic analysis of hybrid renewable energy systems designed for electric vehicle charging: A case study from the United Arab Emirates. *Energies*, 15(18), 6621.
15. Rajagopalan, A. et al. (2024). Multi-objective energy management in a renewable and EV-integrated microgrid using an iterative map-based self-adaptive crystal structure algorithm. *Scientific Reports*, 14(1), 15652.
16. Huang, H. et al. (2024). SREM: Smart renewable energy management scheme with distributed learning and EV network. *Engineering Reports*, 6(5), e12763.
17. Elahe, M.F.; Kabir, M.A.; Mahmud, S.M.H.; and Azim, R. (2023). Factors impacting short-term load forecasting of charging station to electric vehicle. *Electronics*, 12(1), 55.
18. Zhou, D. et al. (2022). Using bayesian deep learning for electric vehicle charging station load forecasting. *Energies*, 15(17), 6195.
19. Xue, M. et al. (2021). Research on load forecasting of charging station based on XGBoost and LSTM model. *Journal of Physics: Conference Series*, 1757(1), 12145.
20. Feng, J.; Chang, X.; Fan, Y.; and Luo, W. (2023). Electric vehicle charging load prediction model considering traffic conditions and temperature. *Processes*, 11(8), 2256.
21. London Borough of Barnet (2024). *Electric Vehicle Charging Transactions*. Retrieved December 12, 2023, from https://www.data.gov.uk/dataset/16c7326b-57fe-4803-88f8-9286c387f68a/electric-vehicle-charging-transactions
22. Vamsi, G. (2024). *Electric Vehicle Charging Stations*. Retrieved December 15, 2023, from https://www.kaggle.com/datasets/gvamsi1999/electric-vehicle-charging-stations
23. Ashvath, S.P. (2023). *Electrical Vehicle Datasets*. Retrieved December 15, 2023, from https://www.kaggle.com/datasets/a2162014/electrical-vehicle-datasets

24. Huang, N.; and He, Q. (2022). *Electric vehicle charging station charging load data set*. Retrieved December 17, 2023, from https://data.mendeley.com/datasets/mgmhz7yfd7/1
25. Shah, S.A.A. (2020). *Past Weather in Kuala Lumpur, Malaysia – January 2020*. Retrieved January 27, 2024, from https://www.timeanddate.com/weather/malaysia/kuala-lumpur/historic?month=1&year=2020
26. Skuza, A.; and Jurecki, R.S. (2022). Analysis of factors affecting the energy consumption of an EV vehicle - A literature study. *IOP Conference Series: Materials Science and Engineering*, 1247(1), 012001.
27. Chapree, C. (2023). *SoyaCincau's Malaysian EV Buyer's Guide – October 2023 Edition*. Retrieved January 8, 2024, from https://soyacincau.com/2023/10/30/soyacincau-malaysian-ev-buyers-guide-oct-2023-edition/
28. Tableau. (2024). *Tableau Prep Builder*. Retrieved April 10, 2024, from https://www.tableau.com/products/prep
29. Tableau. (2024). *Tableau Desktop*. Retrieved April 10, 2024, from https://www.tableau.com/products/desktop
30. Taye, M.M. (2023). Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. *Computers*, 12(5), 91.
31. Dineva, K.; and Atanasova, T. (2020). Systematic look at machine learning algorithms - advantages, disadvantages and practical applications. *International Multidisciplinary Scientific GeoConference*: *SGEM*, 20(2.1), 317-324.
32. Roy, A.; and Chakraborty, S. (2023). Support vector machine in structural reliability analysis: A review. *Reliability Engineering & System Safety*, 233, 109126.
33. *Automodel - Performance*. (2020). Retrieved June 17, 2024, from https://community.rapidminer.com/discussion/57850/automodel-performance
34. Noel_D (2024). *Are There Known Issues with Auto Model's Handling of GBTs?* Retrieved June 26, 2024, from https://community.rapidminer.com/discussion/60683/are-there-known-issues-with-auto-models-handling-of-gbts