PONTIFICAL CATHOLIC UNIVERSITY OF MINAS GERAIS

Graduate Program in Informatics

Marcelo dos Reis

USING ELECTRIC VEHICLE DRIVER'S DRIVING MODE FOR TRIP PLANNING AND ROUTING

Belo Horizonte

2025

Marcelo dos Reis

USING ELECTRIC VEHICLE DRIVER'S DRIVING MODE FOR TRIP PLANNING AND ROUTING

Dissertation presented to the Graduate Program in Informatics at Pontifical Catholic University of Minas Gerais, as a partial requirement to obtain Master's degree in Informatics.

Advisor: Prof. Humberto Torres Marques-Neto

Belo Horizonte

2025

FICHA CATALOGRÁFICA
Elaborada pela Biblioteca da Pontifícia Universidade Católica de Minas Gerais

R375u

Reis, Marcelo dos
Using electric vehicle driver’s driving mode for trip planning and routing / Marcelo dos Reis. Belo Horizonte, 2025.
56 f. : il.

Orientador: Humberto Torres Marques-Neto
Dissertação (Mestrado) - Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Informática

1. Veículos elétricos - Consumo de energia. 2. Aprendizado do computador. 3. Arquitetura de software. 4. Algoritmo. 5. Telemetria. 6. Levantamento de rotas. I. Marques-Neto, Humberto Torres. II. Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Informática. III. Título.

CDU: 681.3.091

Ficha catalográfica elaborada por Gabriela Fernandes dos Santos Dias - CRB 6/4473

Marcelo dos Reis

USING ELECTRIC VEHICLE DRIVER'S DRIVING MODE FOR TRIP PLANNING AND ROUTING

Dissertation presented to the Graduate Program in Informatics as a partial requirement for qualification to the Master degree in Informatics from the Pontifical Catholic University of Minas Gerais

---

Prof. Dr. Humberto Torres Marques-Neto – PUC Minas (Advisor)

Prof. Dr. Felipe Domingos da Cunha – PUC Minas (Examining Board)

Prof. Dr. Vinícius Fernandes Soares Mota – UFES (Examining Board)

Belo Horizonte, August 20, 2025.

Aos meus pais, à minha esposa e a minha família.

AGRADECIMENTOS

Agradeço, primeiramente, à minha esposa Eliane, pelo apoio, incentivo e principalmente paciência.

Agradeço, também, ao meu orientador, Prof. Dr. Humberto, pela confiança depositada, paciência inesgotável, incentivo constante e orientação decisiva. Sempre compreensivo e disposto a ajudar.

Finalmente, agradeço à DMED pela oportunidade de participar dessa pesquisa e pela bolsa de estudos para esse programa de Mestrado.

"To claim absolute knowledge is to become monstrous. Knowledge is an unending adventure at the edge of uncertainty"

Frank Herbert

ABSTRACT

With the increasing adoption of electric vehicles (EVs), challenges such as limited driving range and a scarcity of charging infrastructure have become more pronounced. To address this, we developed and validated ArchDriva, a complete software architecture designed to accurately predict energy consumption and recommend optimal charging stops along a route. Our approach uniquely combines individual driving styles, which we term *drivability*, with trip-context data such as altitude variations, average speed, and auxiliary system usage. This integrated method differentiates our work from studies that typically focus on only one aspect of the problem, and our open architecture is designed for integration with popular navigation applications.

Our methodology is grounded in real-world telemetry data collected from an electric vehicle during October 2023. We enlisted nine drivers to travel a 76-kilometer round-trip route between Poços de Caldas, MG, and São João da Boa Vista, SP, which features a significant altitude variation of approximately 550 meters. Due to a lack of precision in calculating consumption in kWh from the available data, we adopted a strategy based on battery *State-of-Charge* (SOC) percentage, which has a 0.5 pp resolution. To enable a stable estimation of energy consumption, we processed the data using 10 km fixed-distance windows with 99% overlap. The architecture employs K-Means, an unsupervised clustering algorithm, to classify drivers into four distinct classes based on their telemetry patterns (e.g., acceleration and braking). This classification then serves as a key feature, alongside route data, for a regression model tasked with predicting energy consumption. The dataset generated and analyzed during this study is publicly available.

In our evaluation, we tested ten machine learning regression techniques and found that the *Random Forest Regressor* (RFR) delivered the most effective predictions. When using all features, including driver class and trip context, the RFR model achieved a *mean absolute error* (MAE) of 0.249 percentage points. This result is considered highly acceptable, as the error margin is smaller than the battery's measurement granularity of 0.5 pp. Our analysis also revealed a significant discrepancy between estimated travel times from routing APIs and actual travel times, which led us to propose a correction factor, $k_{i,j}$, to adjust the average speed for each trip segment. In conclusion, we successfully validated the feasibility of the proposed architecture, demonstrating that integrating individual driving styles with contextual factors is crucial for developing reliable predictive models for EV battery consumption.

Keywords: Electric Vehicles (EVs), Range Prediction, Energy Consumption Prediction, Machine Learning, Route Planning

RESUMO

A crescente adoção de veículos elétricos (VEs) evidencia desafios significativos que dificultam sua popularização, notadamente a ansiedade de autonomia (medo de a bateria se esgotar) e a infraestrutura de recarga ainda limitada e mal distribuída. Para mitigar esses problemas, este trabalho propõe e valida uma arquitetura de sistema aberta e modular, denominada ArchDriva, projetada para prever o consumo de energia e a autonomia de VEs de forma precisa. O principal diferencial da arquitetura é a integração de múltiplos fatores que influenciam o gasto energético, combinando o estilo de condução do motorista (drivability), dados de telemetria do veículo e informações de contexto da viagem, como variações de altitude, velocidade média, uso de ar-condicionado e modo de economia. O objetivo é fornecer estimativas de consumo por trecho que permitam a recomendação estratégica de pontos de recarga ao longo de uma rota, podendo ser integrada a aplicativos de navegação populares.

A metodologia empregada baseou-se na coleta de dados reais de telemetria de um veículo elétrico, obtidos através da rede CAN e de sensores GPS. O conjunto de dados foi gerado a partir de viagens realizadas por nove motoristas distintos em uma rota de 76 km. A geração e disponibilização pública deste conjunto de dados constituem uma contribuição relevante para a comunidade acadêmica, dada a dificuldade em obter dados reais e detalhados de telemetria de veículos elétricos, que frequentemente são restritos a sistemas proprietários de montadoras. Para a análise, os dados foram segmentados em janelas de 10 km com 99% de sobreposição, permitindo uma avaliação granular do consumo percentual da bateria. A arquitetura implementa um classificador não supervisionado, utilizando o algoritmo K-Means, para agrupar os motoristas em quatro classes distintas de acordo com seus padrões de aceleração, frenagem e velocidade. Subsequentemente, um modelo de regressão é treinado para prever o consumo de bateria, utilizando como atributos a classe do motorista e as características de cada trecho da viagem.

A avaliação da arquitetura demonstrou sua viabilidade e eficácia. Foram testadas dez técnicas de aprendizado de máquina para a tarefa de regressão, e o algoritmo Random Forest Regressor apresentou o melhor desempenho, com um erro médio absoluto de 0,249 pontos percentuais (pp), valor considerado satisfatório frente à resolução de 0,5 pp do medidor de bateria do veículo. A análise também revelou que a inclusão dos dados de contexto da viagem e da classificação do motorista reduz significativamente o erro de previsão. Um desafio identificado foi a discrepância entre os tempos de viagem estimados por APIs de roteirização e os tempos reais, o que levou à proposição de um fator de correção para a velocidade média de cada trecho. Conclui-se que a arquitetura ArchDriva é uma solução robusta e precisa para a previsão de autonomia, reforçando a importância de considerar o comportamento individual do motorista e as condições da via para otimizar o planejamento de viagens em veículos elétricos.

Palavras-chave: Veículos Elétricos (VEs), Previsão de Autonomia, Previsão de Consumo de Energia, Aprendizado de Máquina, Planejamento de Rota

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

API – *Application Programming Interface*

CaaS – *Containers as a Service*

CAN – *Controller Area Network*

DTR – *Decision Tree Regressor*

ENR – *Elastic Net Regression*

FaaS – *Function as a Service*

FMM – *Fuzzy Markov Model*

GBR – *Gradient Boosting Regressor*

GPRS – *General Packet Radio System*

GSM – *Global System for Mobile Telecommunications*

HMM – *Hidden Markov Model*

IaaS – *Infrastructure as a Service*

KNR – *K-Neighbors Regressor*

LR – *Linear Regression*

MAE – *Medium Average Error*

MQTT – *Message Queue Telemetry Transport*

OBD-II – *On Board Diagnostics II*

PaaS – *Platform as a Service*

RFR – *Random Forest Regressor*

RMSE – *Root Mean-Squared Error*

SOC – *State of Charge*

SVR – *Support Vector Regression*

CONTENTS

1 INTRODUCTION

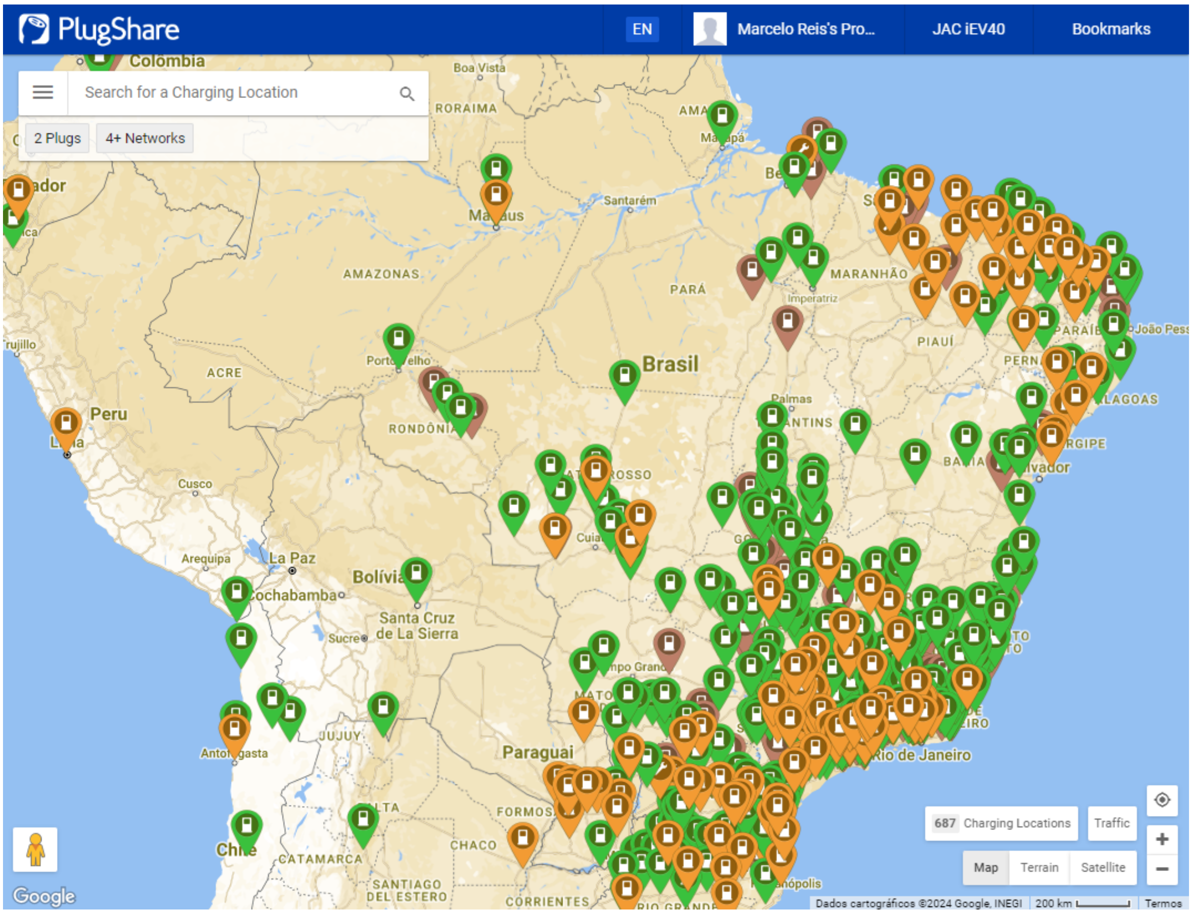
Projections under the Stated Policies Scenario (STEPS) indicate substantial growth in the electric vehicle (EV) market, with the global fleet of EVs, excluding two- and three-wheelers, expected to reach 250 million units by 2030. This expansion reflects an average annual growth rate of approximately 25% and represents a four-fold increase from the end of 2024, resulting in EVs constituting approximately 15% of the total global vehicle fleet when all vehicle types are included (International Energy Agency, 2025). However, even with tax over automotive vehicles exemption or reduction in some countries (like Brazil) and discount on the import tax of electric vehicles, largely used strategies to accelerate the adoption (ALI; NAUSHAD, 2022), the high cost of acquisition has been preventing its widespread adoption (VARGAS et al., 2020). Companies such as Ambev, Coca-Cola and JBS have been using the distribution fleet electrification to achieve carbon footprint reduction goals. It is expected that by 2040, about 85% of the fleet of cars per application will be electric (MCKINSEY, 2023). Reducing the negative environmental impact per capita of cities is a key objective of SDG 11 (Sustainable Development Goal 11 - Sustainable cities and communities) and the recharge planning of electric vehicles in long-distance trips can be an important factor to encourage more people to adopt electric vehicles.

Another common problem with the large-scale adoption of electric vehicles is the recharges and low autonomy problems (LEBROUHI et al., 2021), such as: (i) high frequency of recharges, due to the limited capacity of the batteries; (ii) long wait to recharge, typically 30 minutes to 2 hours; (iii) recharge peaks, as there are more popular times to recharge, such as lunch breaks and others. In Brazil, it is also important to highlight the limited supply of public electric stations on roads and cities, as shown in Figure 1. In part, it happens due to the fact that only in June 2018 ANEEL* published a resolution giving permission to commercial exploitation of electric vehicle supply (VARGAS et al., 2020). In other countries, similar issues with the charging infrastructure exists, as we can observe in Italy, which lacks investments in Southern regions, shown in Figure 2. Turkey's deficiencies in the existing charging infrastructure is illustrated in the density map in Figure 3, and in Figure 4 we can see the same problem of poor distribution of stations in Australia, notably between Adelaide and Perth.

For these reasons, the correct autonomy prediction and the consequent recommendation of charging stations along the way is one of the approaches to partially solve some

*Brazilian National Electric Energy Regulator.

Figure 1 – EV charging infrastructure in Brazil in April, 2024

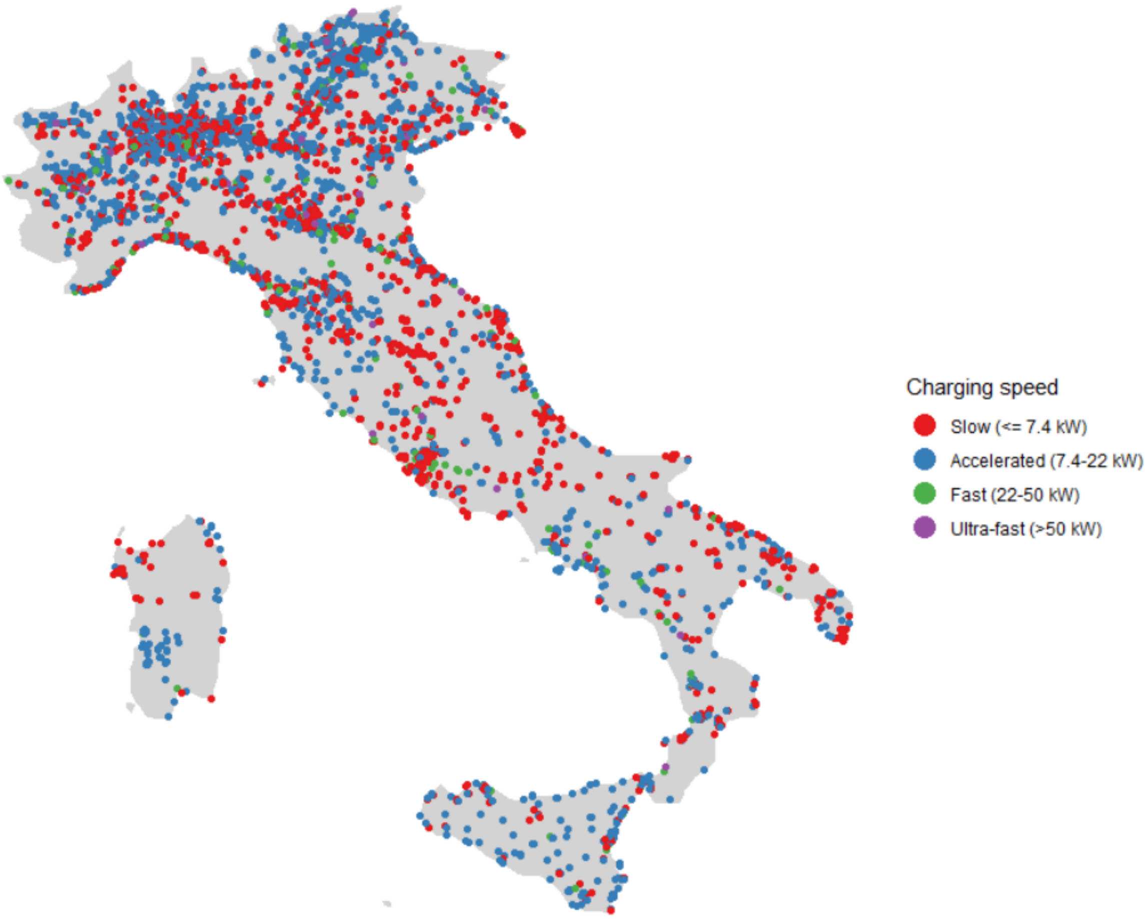


Source: Available at `https://www.plugshare.com`. Accessed on 04/02/2024

of these problems (ALANAZI, 2023). Indeed, one of the most important factor that can influence the autonomy prediction is the driving manner (Delnevo et al., 2019). How a person drives a vehicle is a complex concept that can be defined as the way the driver operates the vehicle's control in the context of driving and external conditions. The characterization of the way of driving can be decisive for a series of situations, such as predicting dangerous behaviors, avoiding traffic accidents, and others (Fugiglando et al., 2019). From now on, we call way of driving as *drivability*.

There are works predicting the consumption or autonomy of electric vehicles using telemetry data and physics formulation (BAILEY et al., 2022), regression algorithms (CHANDRAN et al., 2021) or deep neural network (ADEDEJI; KABIR, 2023a). Other articles focuses on the driving mode versus fuel economy relation (Stanton; Allison, 2020), while several papers studies diverse ways of driving (Siami; Naderpour; Lu, 2020). Also, Bustos et al. (2025) introduced the concept of Maximum Driving Range to predict the worst case scenario in a planned route. More recent works, presents an Hybrid DNN

Figure 2 – EV charging stations' locations in Italy

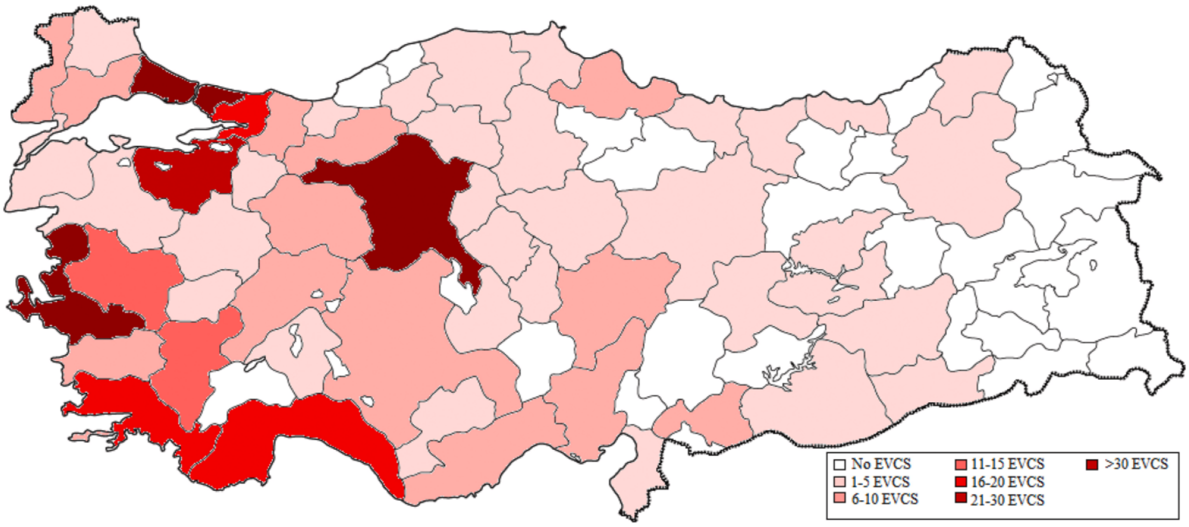


Source: Noussan (2020), Fig. 2

method (MUSTAFFA; SULAIMAN; ISUWA, 2025).

In this paper, we propose a system architecture that makes use of a motorist type classifier and a model for autonomy prediction. We use telemetry data to determine and classify drivers by their driving style and to forecast the energy consumption of each leg of a trip. Our goal is to make the prediction considering the type of driver and the context data of the segment. In this way, it is possible to recommend charging stations along the route, considering the calculated ranges, to offer a better experience to the driver. Although there are proprietary applications, such as Tesla's and Volvo's, this work describes an open architecture, with *Application Programming Interface* (API) that can be used by more popular vehicles and integrated with existing applications, such as *Waze* and *Google Maps*.

Figure 3 – EV charging stations density map of Turkey



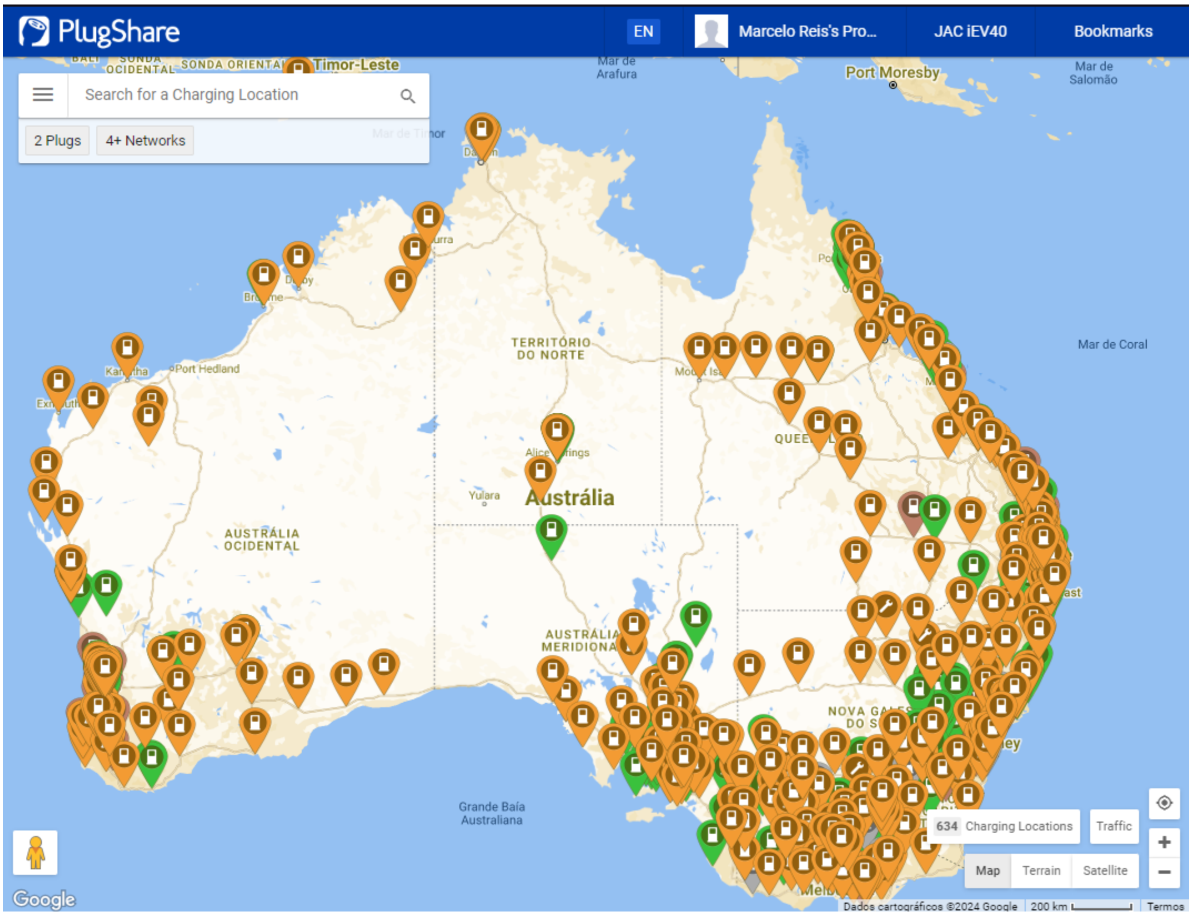
Source: Gönül, Duman e Güler (2021), Fig. 8

Thus, we propose a system capable of predicting the autonomy of electric vehicles and allow recommendation of charging points. Moreover, (a) we assess whether the proposed architecture meets the purpose of locating charging stations along a trip; (b) we identify a very suitable Machine Learning algorithms to predict the range of electric vehicles; and (c) we incorporate the *drivability* and travel context evidence into selected recommendation strategies. This approach differentiates our method from other studies that typically address only individual aspects of *State-of-Charge* (SOC) prediction, consumption forecasting, or driving pattern analysis.

After analyzing real data collected directly from an electric vehicle, we used k-means to classify the drivers and also verified some regression methods that obtained satisfactory results to predict the amount of battery energy consumed in travel segments. The *Random Forest Regressor* (RFR) method obtained an absolute average error of 0.249 pp. Therefore, through the routing application, it is possible to predict the autonomy of a projected trip, using the consumption forecast section by section. In addition, based on the calculated range and the route traced, we propose to recommend charging stations.

This master thesis is organized as follow. Chapter 2 explores the concepts ne-

Figure 4 – EV charging infrastructure in Australia in April, 2024



Source: Available at `https://www.plugshare.com`. Accessed on 04/02/2024

cessary to understand the designed architecture, as well as the modules introduced to address the problem. We then survey existing studies on driving-mode classification and electric-vehicle range prediction in Chapter 3. The design of our system architecture appears in Chapter 4, which also details our methodology for data collection and preprocessing. Next, Chapter 5 presents the main evaluation results. Finally, we summarize key findings and discuss future directions in the Chapter 6.

2 THEORETICAL FOUNDATIONS

For the proposed architecture, we use several technologies integrated into modules and we use telemetry data in classification and regression algorithms to predict battery consumption and autonomy. The main technologies are described below.

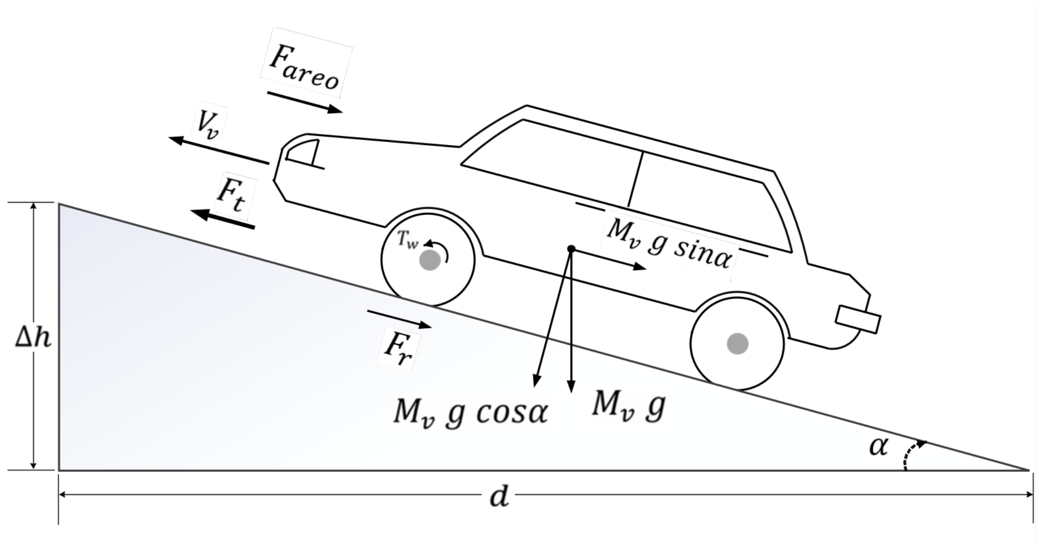
2.1 Physical Modeling of Energy Consumption

Akl, Ahmed e Rashad (2019) describes the forces acting on a moving vehicle, as shown in Figure 5. Thus, the force required to propel a vehicle at a given moment is calculated as:

$$F_t = F_{aero} + F_r + M_v g \sin \alpha \tag{2.1}$$

where M_v, g, and α represent the vehicle mass, gravitational acceleration, and road incline, respectively. The drag force, F_{aero}, depends on air resistance and correlates with speed (HUCHO, 1998), while the gravitational force to overcome depends on the slope.

Figure 5 – Forces acting on a vehicle ascending a slope



Source: Akl, Ahmed e Rashad (2019)

Additionally, electric vehicles include an economy mode, which functions inversely to the sport mode in automatic transmission combustion cars, limiting energy consumption when activated. Among auxiliary systems, the air conditioning system is the primary energy consumer, drawing power directly from the battery. Energy consumption follows:

$$E_s = E_t + E_a \tag{2.2}$$

where E_t and E_a represent the energy required to generate F_t over time and the energy consumed by auxiliary systems in the same period. Based on this, the most influential features for energy calculation must be selected.

2.2 Vehicle Telemetry via CAN Bus

With the increase in the fleet, the usage of telemetry data from electric vehicles can be an important source of data to determine the way of driving, to predict how energy is expended and, therefore, the vehicle's autonomy. The *Controller Area Network* (CAN) is a high-speed serial bus system common in embedded systems (ISO Central Secretary, 2015). It is a technology widely used in modern vehicles from the adoption of various micro-controllers, such as electronic injection, ABS and others. Through this bus, the data of acceleration, speed, odometer, pressure on the accelerator pedal and other data of electric vehicles are obtained (FUGIGLANDO et al., 2018).

This data can be transmitted over the Internet, through a communication protocol. With the increase in the usage of cell phones, the need for standardization of communication protocols arose, culminating in the specification of *Global System for Mobile Telecommunications* (GSM). Furthermore, with the need for data transmission and access to the Internet, the *General Packet Radio System* (GPRS) protocol arises, specifically for connecting to the Internet (HALONEN; ROMERO; MELERO, 2004). By using GPRS, available in 3G and 4G coverage areas, it is possible to cover almost all Brazilian cities and 44.2%, which corresponds to 55,711 km, of their highways, according to ANATEL*'s infrastructure data panel (ANATEL, 2022).

2.3 Drivability: Characterizing Driving Style

Characterizing driving behavior is complex and depends on contextual and external factors. However, telemetry data, obtained via CAN, can identify drivers with minimal interference from these factors. Fugiglando et al. (2019) analyzed eight CAN bus signals that are related to how the driver operates the vehicle. Pedals and steering-wheel signals are directly responsible for vehicle response, and others, like speed or accelerations, represents the feelings or habits of the conductor.

Driving style may significantly impact vehicle fuel consumption and likely affects

*Brazilian National Telecommunication Regulator.

the battery consumption of electric vehicles. In Ping et al. (2019), some of these signals have been explored for their correlation to fuel consumption, in particular, gas/accelerator and brake pedal position, steering angle, speed and acceleration.

Using these features, we aim to characterize driving style, cluster drivers with similar behavior, and simplify our regression model.

2.4 Pearson Correlation Analysis

The Pearson coefficient measures the strength and direction of the linear correlation between two variables. Represented by ρ, it ranges from -1 to 1 (SOUSA, 2019). The correlation can be:

- $0.9 \leq |\rho| \leq 1.0$: very strong correlation;
- $0.7 \leq |\rho| < 0.9$: strong correlation;
- $0.5 \leq |\rho| < 0.7$: moderate correlation;
- $0.3 \leq |\rho| < 0.5$: weak correlation;
- $0.0 \leq |\rho| < 0.3$: no correlation.

2.5 Unsupervised Classification with K-Means

To classify drivers by driving mode, we propose an unsupervised method of clustering data analysis. Using unsupervised clustering methods, we can determine which parameters are more suitable to identify groups from the analyzed data (ARABIE; HUBERT; SOETE, 1996).

The accuracy of clustering methods deeply depends on the number of groups and features. In addition, the number of analyzed features influences in the groups characterization, increasing or decreasing the number of needed classes for differentiation. Also, having few groups may result in very generic groups with few similarities, but many groups would make the model very complex and computationally expensive. Finding the right number of groups and features, without losing information, has already been the subject of study (Fugiglando et al., 2019; RODRIGUEZ et al., 2019) and is one of future goals.

2.6 Regression Models for Prediction

We use linear regression algorithms to determine the behavior of a dataset and extrapolate output values based on the input data. We seek to find a correlation between the independent variables and the results. Multiple linear regression algorithms are an extension of the simple linear regression model, where multiple prediction variables are used for only one outcome of a continuous set (EBERLY, 2007).

To forecast of consumption, we propose to use a multiple regression algorithm, using travel context data, section by section, with its consumption and type of driver. And, based on this information, predict the consumption and autonomy, for that type of driver, for the next calculated stretches.

2.7 Evaluation Metrics for Regression

Assessing predictive model performance requires evaluation metrics, with *Root Mean-Squared Error* (RMSE) and *Medium Average Error* (MAE) being the most widely used, though there is no consensus on which is superior (CHAI; DRAXLER, 2014).

The RMSE metric is based on the squared mean differences between predictions and actual observations, calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.3}$$

MAE is the simplest metric to interpret, calculated by computing the absolute difference for each point, then averaging these differences:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2.4}$$

2.8 Statistical Validation with Bootstrap

The *bootstrap* method is commonly used to estimate unknown populations from sample data and serves as a crucial tool, particularly in complex models (BOOS, 2003). In this study, we performed 1,000 bootstrap resamplings to measure the population's average error.

2.9 Scalability in Cloud Architectures

A system's ability to scale according to workload variations is increasingly crucial. System scalability is measured in three dimensions, according to Brataas and Hughes (BRATAAS; HUGHES, 2004):

(i) Processing capacity: speed of task execution

(ii) Information capacity: amount of stored data

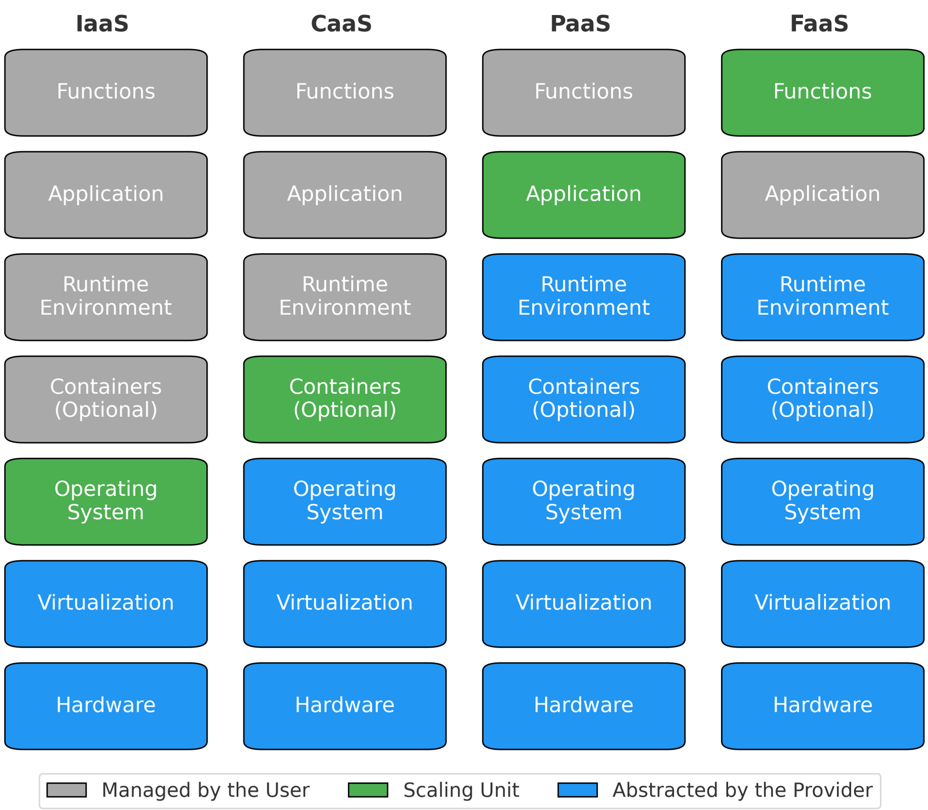
(iii) Connectivity: number of simultaneous connections

A scalable architecture can increase these dimensions proportionally to the system's needs.

With the advent of cloud computing, various architectures support scalable systems, as illustrated in Figure 6:

- *Infrastructure as a Service* (IaaS): scalable virtual machines;
- *Containers as a Service* (CaaS): containerization using *Kubernetes* or *Docker Swarm*;
- *Platform as a Service* (PaaS): specific services such as databases, file storage, or execution environments;
- *Function as a Service* (FaaS): also known as microservices or *Serverless*.

Figure 6 – Infrastructure abstraction levels



Source: The authors

3 RELATED WORK

A widely cited barrier to the broader adoption of electric vehicles is range anxiety—the concern that the battery will be exhausted before reaching a destination or an appropriate charging station. Reducing this concern has motivated substantial research on frameworks for estimating an EV's remaining range and *State of Charge* (SOC) (PRASAD; GUDIPALLI, 2025). At the core of this problem is the prediction of energy consumption, which is influenced by multiple interacting variables.

The first line of work in this area is vehicle-centric. Many early and current models estimate SOC from battery-related features using machine learning (CHANDRAN et al., 2021). To improve these estimates, subsequent studies have incorporated additional vehicle telemetry, such as instantaneous speed and mileage, as inputs to deep neural networks (ADEDEJI; KABIR, 2023a). This direction has produced a range of models, including feedforward DNNs for accurate SOC prediction (ADEDEJI; KABIR, 2023b) and multifunctional ANNs for broader consumption forecasting (ADEDEJI, 2023a).

Recognizing that vehicle data alone is not sufficient, a second line of work adds external context. Route topography—especially altitude variation—has been identified as a key determinant of energy use and effective range (BOLOVINOU et al., 2014; WANG et al., 2024). Other studies include roadway characteristics and auxiliary loads (e.g., air conditioning) to refine SOC estimation (TRAN et al., 2023). Environmental conditions have also been examined, with wind and temperature highlighted as important variables in comparative analyses of AI models (PAUL et al., 2023a).

Beyond the vehicle and its environment, driver behavior (*drivability*) is a major source of variability in consumption. A substantial literature links driving style to fuel economy in general (Stanton; Allison, 2020; Ping et al., 2019), with specific attention to battery savings in EVs. Models based on LSTMs (Huang et al., 2020) and neural networks (Chen et al., 2019b) explicitly relate driving behavior to battery usage, including the effects of particular driving patterns (TIWARY; GARG; MISHRA, 2022) and the role of eco-driving as a means to reduce consumption (GOUJON; HAGHIGHI; HAMRI, 2022).

Given the importance of drivability, its identification and classification form an active research area. Most classification models rely on high-frequency data from CAN or

Table 1 – Comparative analysis of recent works on SOC (State of Charge) prediction

Paper Reference	Telemetry data	Trip context data	Drivability	Consumption prediction	Method	Route planning
Adedeji e Kabir (2023a)	yes	no	no	yes	ANN	no
Tran et al. (2023)	yes	yes	no	yes	Mathematical	yes
R., Venkateshkumar e Kuppusamy (2023)	yes	no	no	yes	LTSM	no
Adedeji (2023b)	yes	no	no	yes	ANN	no
Liu, Shi e Fan (2022)	yes	yes	yes	yes	ANN	no
Mashkov e Karova (2023a)	yes	no	no	yes	Mathematical	no
Kocaarslan et al. (2022)	yes	yes	yes	yes	Mathematical	no
Tiwary, Garg e Mishra (2022)	yes*	no	yes	no	-	no
Goujon, Haghighi e Hamri (2022)	yes	yes	yes	yes	Mathematical	no
Wei et al. (2022)	yes	yes	no	yes	Mathematical	no
Paul et al. (2023b)	yes	yes	no	yes	LTSM, DNN, CNN and Regression Models	no
Bailey et al. (2022)	yes	yes	no	yes	Mathematical	yes
Wang et al. (2024)	no	yes	no	yes	Regression Models	no
Bustos et al. (2025)	no	yes	no	yes	Regression and Physics	no
Mustaffa, Sulaiman e Isuwa (2025)	no	yes	yes	yes	Hybrid DNN	yes
this paper	yes	yes	yes	yes	Regression Models	yes

On Board Diagnostics II (OBD-II) networks (Fugiglando et al., 2019; Chen; Lin; Chen, 2019; Chen et al., 2019a; Ping et al., 2019; Krishnamurthy et al., 2019), with some studies reporting dozens of distinct styles (Siami; Naderpour; Lu, 2020). A variety of methods has been used, including deep learning (Sama et al., 2020), convolutional and recurrent neural networks (Jia et al., 2020; Li; Kang, 2020), *Hidden Markov Model* (HMM) and *Fuzzy Markov Model* (FMM) (JIANG et al., 2025), and combinations such as Kohonen maps with autoencoders (Siami; Naderpour; Lu, 2020; Jia et al., 2020). Unsupervised approaches like k-means are frequently adopted to group drivers by habits (Fugiglando et al., 2019; Warren; Lipkowitz; Sokolov, 2019; Liu; Wang; Wang, 2019).

While accurate prediction is essential, its practical value appears in route planning. More recent and integrated studies aim to close this gap. Some works predict remaining range in complex urban settings with charging and discharging events (JIANG et al., 2025); others employ variable-parameter modeling (MASHKOV; KAROVA, 2023b), AMRIT-based prognostics (R.; VENKATESHKUMAR; KUPPUSAMY, 2023), or federated learning (THORGEIRSSON et al., 2021) to build more resilient frameworks. Concepts such as *Maximum Driving Range* have also been proposed to address worst-case scenarios and further reduce range anxiety (BUSTOS et al., 2025).

However, as summarized in Table 1, most studies address only part of the problem. Although the literature offers strong models for consumption prediction and for driver classification, there is still a need for a *complete and modular software architecture* that integrates these elements into an end-to-end route-planning solution. None of the reviewed works presents an open architecture that combines vehicle telemetry, external trip context, and a distinct driver-classification module specifically to recommend charging stops, as we address by presenting and validating *ArchDriva*.

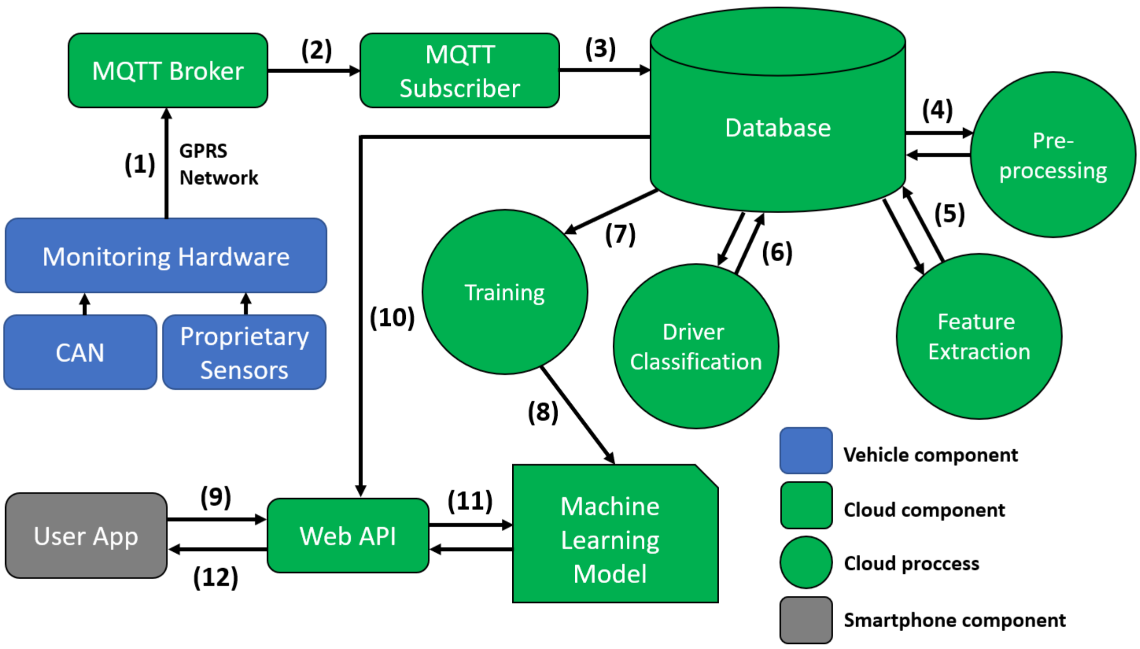
In our previous work (REIS; TEIXEIRA; MARQUES-NETO, 2023) (in Portuguese), we presented an architecture that integrates: (a) a driver classification algorithm dividing them by driving mode, using features extracted from telemetry data, with (b) a linear regression algorithm to determine consumption section by section, taking into account the type of driver and the trip context data, to determine (c) the range and recommend charging stations along the journey.

In this work, we extend the results by testing with a dataset with more drivers, each completing a single session along the same round-trip route under comparable operating conditions, while retaining the original acquisition protocol and preprocessing steps. Feature extraction follows the same specification, with ACC, BRK, and SPD among the core signals. Holding the route and topography fixed reduces contextual confounders and enables direct between-driver comparisons, while the larger group strengthens external validity and supports a stricter evaluation of generalization.

4 PROPOSED ARCHITECTURE AND METHODOLOGY

Our proposed architecture comprises multiple integrated components and processes collaborating to provide charging station recommendations. These components include both in-vehicle elements and cloud-based functionalities, as illustrated in Figure 7, with arrows denoting the flow of information between them. Importantly, all these components are designed to be modular, ensuring ease of replacement as required.

Figure 7 – Proposed system architecture - ArchDriva



Source: The authors

Figure 7 also shows the monitoring hardware consisting of an equipment developed by a research team of PUC Minas (our University) to capture, through the CAN network and its own sensors, the necessary telemetry data. This equipment collects the data and sends it to a *Message Queue Telemetry Transport* (MQTT) Broker, through a GPRS connection (1), over the Internet. The MQTT broker receives this telemetry data (1) and transmits it to its subscribers (2). The MQTT subscriber receives messages warning of the data change and reads it (2), writing it to a Database (3). In the pre-processing module, the collected data, read from the database (4), identified with the respective drivers, are grouped, normalized, re-sampled and completed.

The features are extracted in the feature extraction process, consisting of two sets: (i) features to classify drivers, like (a) throttle and break pedal pressure; (b) speed; (c) inclination; (d) acceleration; and (ii) features to linear regression, which consists of analyzing the data (5) of each trip, extracting, in ten-kilometer windows: (a) vehicle; (b) driver; (c) average speed; (d) altitude variation; (e) use of air conditioning; (f) economy mode; (g) consumption. Also shown in Figure 7, the driver classification process reads data from table of features (6), limited to a maximum period of one year and, using K-Means algorithm, groups drivers into driver's classes. In the training stage, the regression model is trained with data from the features table, through data-flow (7), to forecast consumption, generating and saving a model (8), which calculates the energy consumption in that stretch, according to the type of driver and the data of the trip context, such as altitude variation, speed, air conditioning usage and economy mode.

The Machine Learning Model is called directly by the API and receives: (a) driver identification and (b) trip data, section by section, and returns all sections to the API, with their respective consumption. This Web API is used by a mobile application that traces the route according to the user's starting point and the chosen destination. This application can use one route API to determine the data from the stretches as accurately and with the greatest possible sampling. This API receives route information and driver identification (9). The driver's class is determined by consulting the table in the database (10), according to the classification of drivers. Then, the model is called with the driver classification and trip data, returning the consumption estimate section by section (11). The User Application is responsible for connecting the driver to the vehicle and its main function is to trace the travel route and recommend stops for recharging along the way.

4.1 Data Collection and Preparation Methodology

To evaluate this architecture, we used the telemetry data from a real electric vehicle. These data were collected in October 2023, through an ANEEL R&D, which acquired the vehicle for DME Distribuição S.A. (DMED). We developed a device to capture, through the CAN network, the following telemetry data: (a) Instantaneous speed; (b) State of charge - % of electric vehicle battery; (c) Odometer; (d) Engine voltage; (e) Engine current; (f) Tilt of the accelerator pedal; (g) Brake pedal tilt; (h) Air conditioning (on/off); (i) Economy mode (on/off); (j) Expected autonomy.

In addition to these data, the equipment also provides **GPS!** (**GPS!**) data, such as: (a) Positioning (latitude and longitude); (b) Electric vehicle tilt (gyroscope); (c) Acceleration (accelerometer); (d) Altitude; (e) Date; (f) Time. This equipment collects the data and sends it to an MQTT Broker, through a GPRS connection, that reads and saves to a database. The collection routine continuously generates data at a frequency of two or three seconds between each collection.

Since the sampling interval ranged between 2 and 3 seconds, we adopted an strategy using the battery discharge expressed as a percentage. This choice was motivated by the availability of SOC values reported in discrete intervals, with a resolution of 0.5 pp.

Data pre-processing is performed once a day, analyzing the data stored from the previous day and saving the processed data to the database. For this processing, some considerations were made:

1. each trip consists of a collection of chronologically sequential data, from the same vehicle and the same driver on the same day.
2. the beginning and end of each trip are marked when the speed is zero and the brake and accelerator pedal are not pressed, indicating car at rest.
3. when a trip comprises more than one day, it is separated into more than one trip.
4. every one-minute period with no data, a new trip is also considered, as it is not feasible to check the missing values.

4.2 The Experimental Dataset

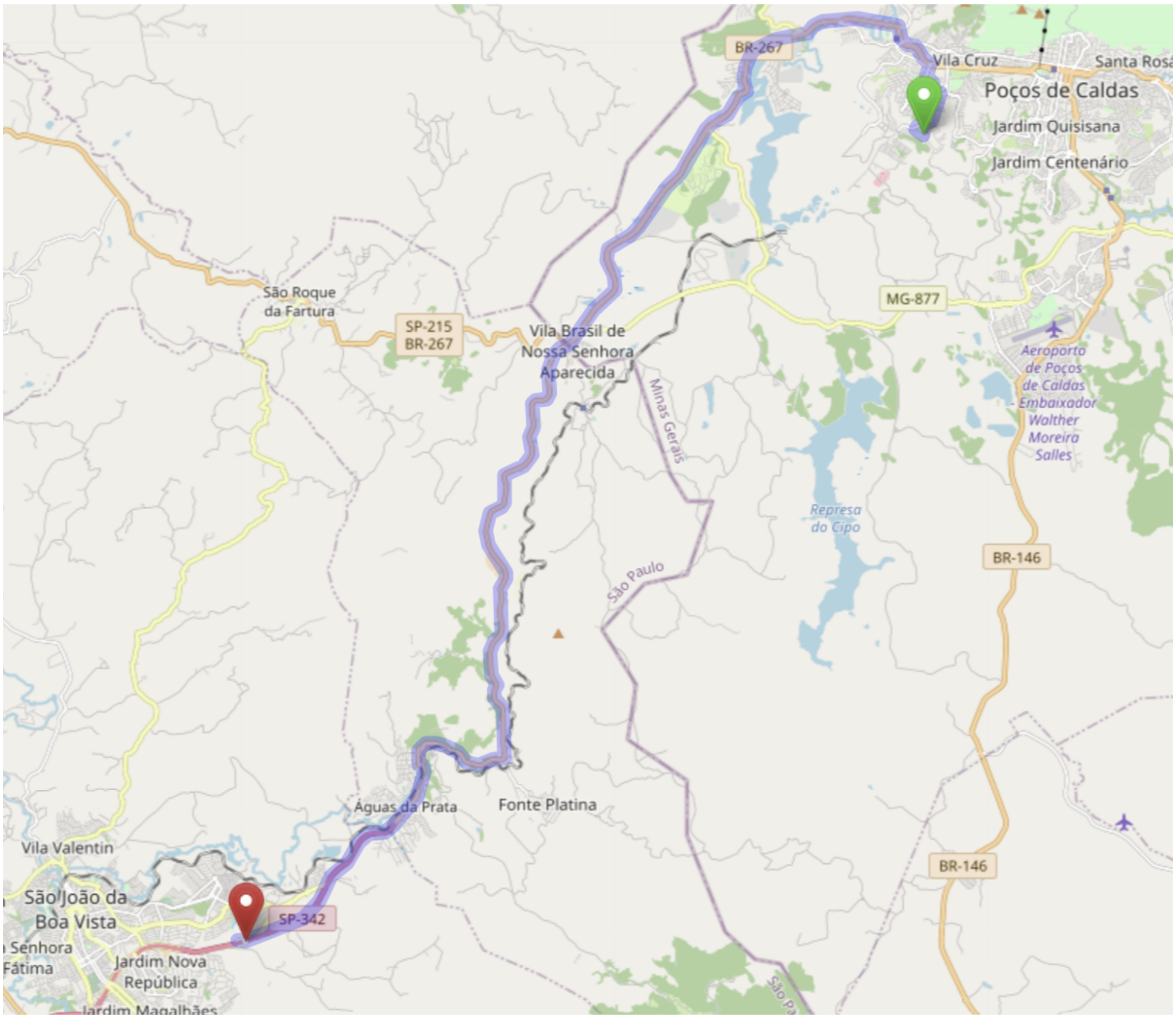
In our data collection study, we enlisted the participation of nine drivers to traverse the same round-trip route connecting Poços de Caldas, MG, and São João da Boa Vista, SP. This route spans approximately 76 kilometers in length and exhibits a notable altitude variation of approximately 550 meters, visually represented in Figure 8, using OpenStreetMap*.

Thus, the database is composed of 24,277 telemetry records, collected in October 2023. After pre-processing, the resulting dataset consisted of 24,113 records, with 696.4 km of travel recorded, divided in 9 identified trips, as detailed in Table 2. Table 3 show the dataset summarized by the economy mode and the air conditioning status.

However, we observed that evaluating battery discharge over very short distances, such as 1 km, 2 km, or even 5 km, produced variations that were either negligible or too noisy to reflect meaningful consumption patterns. This was due to the limited granularity of the SOC readings, which did not change consistently over such short intervals, as in Table 4. For the 3 km window, the mean consumption is 0.52%. Given the sensor resolution of 0.5%, this means an average trip segment registers only a single tick of change in the battery's state of charge. A significant number of segments (where consumption is $< 0.5\%$) will have a recorded CONS of 0.0%, while others will be quantized to 0.5%.

*Available at https://www.openstreetmap.org/

Figure 8 – Travel round-trip route - 76 km - OpenStreetMap



Source: The authors

The target variable for this dataset is therefore extremely coarse and suffers from high quantization noise. For the 10 km window, the mean consumption is 1.75%. This means an average segment would register 3 or 4 discrete steps from the sensor (i.e., a drop from 80.0% to 78.5% or 78.0%). While still discrete, this provides a much higher-resolution signal. The quantization error is a smaller fraction of the overall measurement, making CONS a more reliable target variable.

To address this limitation, we defined fixed-distance windows of 10 km for our analysis. This distance was sufficient to capture measurable variations in SOC, enabling a more stable estimation of average energy consumption per kilometer. By aggregating data over larger segments, we reduced the impact of individual measurement noise and improved the reliability of our consumption estimates. With data from ten-kilometer

Table 2 – Trips after pre-processing

Driver	Distance	Battery Discharge	Average Speed
#1	76.4 km	36.0%	57.96 km/h
#2	75.8 km	37.5%	63.46 km/h
#3	76.3 km	46.0%	64.07 km/h
#4	76.0 km	35.5%	57.07 km/h
#5	68.9 km	24.0%	55.03 km/h
#6	76.0 km	33.0%	53.08 km/h
#7	76.2 km	38.0%	53.92 km/h
#8	76.2 km	35.5%	53.62 km/h
#9	19.3 km	12.5%	55.23 km/h

Table 3 – Dataset after pre-processing

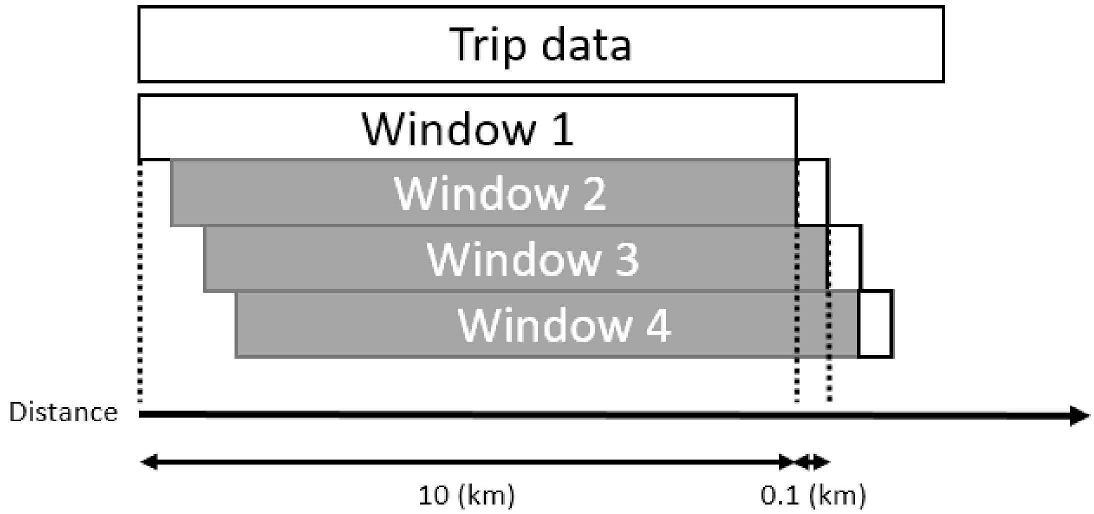
Economy Mode	Air Conditioning	Number of Records
on	on	364
	off	721
off	on	10,518
	off	12,510

stretches, the feature extraction phase extracted 5,285 records. To extract the stretches and maximize the number of records to improve the regression model, we use windows of 10 km for each 0.1 km variation of the same trip, overlapping 99% of each window, as shown in Figure 9. We have chosen 10 km for the window size to have relevant battery discharge and 0.1 km because that is the minimum step in the car odometer available from sensor.

Table 4 – Analysis of window sizes

Window Size	Number of Instances	Mean Consumption	Std Dev Consumption
3 km	2641	0.52%	0.23%
4 km	1978	0.70%	0.31%
5 km	1579	0.87%	0.39%
6 km	1311	1.05%	0.47%
7 km	1118	1.22%	0.55%
8 km	973	1.40%	0.63%
9 km	858	1.57%	0.71%
10 km	780	1.75%	0.79%

Figure 9 – Moving window with 10 km length and 99% overlap



Source: The authors

4.3 Architectural Modules

4.3.1 Driver Classification Module

With the analysis of this data, we verify the best features to identify driving patterns, with respect to instantaneous battery consumption. Some metrics used in the classification of drivers are acceleration, inclination, speed, and pedal position, among others. In order to identify these patterns, as it is not possible to define the number of classes or which drivers will be of each class a priori, we used an unsupervised classification learning techniques, K-means. In this paper, we used 4 classes and the following features:

- **AX**, **AY** and **AZ**: the acceleration in each axis, given in g-force;
- **GX**, **GY** and **GZ**: the inclination in each axis, in º/s;
- **BRK**: brake pedal position, as % of pedal travel;
- **ACC**: accelerator (throttle) pedal position, as % of pedal travel;
- **SPD**: instantaneous velocity, in meters per second.

It consists of reading the data from table (6) of features, for a maximum period of one year, generating histograms of 10 equally divided bins, for each driver and vehicle. We used the conversion of values into histograms to reduce the number of values of each feature, aiming to simplify the processing of the classification model, without losing important information, as in Fugiglando et al. (2019). Using the histograms of each driver as input, an unsupervised classifier groups the drivers into classes, saving the driver and class tuples in a database table (6). According to Alpaydin (2020), one of the most used classification algorithms is K-Means, which is an unsupervised method of classification and grouping of data that analyzes the distribution of data and identifies centroids to generate groups with similar characteristics. Typically, K-Means input parameters are: an integer k, representing the number of classes, and a sample $x = x_1, x_2, ..., x_n$. Thus, the clusters are decided by randomly generating centroids and choosing the closest centroid for each x_i. Next, the center of each cluster is calculated and it is verified whether each object remains closer to these new centroids, repeating these steps until there is convergence. Each element in the sample can have one or more features that characterize the element.

4.3.2 Consumption Prediction and Training Module

We created another set of features for use in estimating consumption, based on the type of driver and the trip context data. In Akl, Ahmed e Rashad (2019), the forces acting on a moving car are described. Thus, the force necessary to propel the car at a given moment is calculated by

$$F_t = F_{aero} + F_r + M_v g \sin\alpha \tag{4.1}$$

where; M_v, g and α are the vehicle mass, gravitational acceleration and the slope of the road, respectively. The air drag resistance force, F_{aero}, is correlated to speed (HUCHO, 1998) and the gravity force applied depends on the inclination of the slope. In addition, electric vehicles have an economy mode, which works opposite to sport mode on regular cars, spending less energy when switched on. Besides that, the main auxiliary energy burner is the air conditioning system, which drains energy directly from the battery. The energy is drained a in

$$E_s = E_t + E_a \tag{4.2}$$

where; E_t and E_a is the energy necessary to generate F_t through a given time (E_t) and the energy spent on auxiliary systems on the same time (E_a). Based on that, we selected some features that might have influence on energy consumption. These features are described below, along with how they are calculated, for each ten-kilometer stretch:

- **DIST**: travelled distance in kilometers;
- **VEIC**: vehicle identification - each vehicle has its characteristics and battery charge capacity;
- **COND**: identification of the driver's class - for each class, there is a way to drive, more or less economical, depending on the route data;

- **VARALT**: altitude variation in the section in meter per meter - calculated as the difference in altitude at the start and end of the section divided by its size in meters;
- **SPD**: average speed in the segment, in meters per second - calculated as the distance covered in the segment by the time spent;
- **ARC**: use of air conditioning in the section - 0 for off and 1 for on - intermediate values indicate the percentage of use in the section;
- **ECO**: use of economy mode in the section - 0 for off and 1 for on - intermediate values indicate the percentage of use in the section;
- **CONS**: percentage of battery consumed in the section, with an accuracy of 0.5 pp, as informed by the vehicle.

Using the data from the features table (7), replacing the driver's identification for his class, a model is generated, through multiple linear regression, to forecast consumption, section by section. This model analyzes, according to the type of driver and the data of the trip context, such as altitude variation, speed, air conditioning use and economy mode, the percentage of battery used in that segment. Adding up the cost section by section, we have the percentage value of battery used in the total journey. This model is saved as a file in the cloud (8) for use in forecasting consumption section by section and this training can be carried out daily or at longer intervals according to the available computational resources.

The model is called directly by the API (11) and receives:

- The driver's identification
- Trip data, leg by leg

This module loads the linear regression model saved from the "Training" module and, for each segment of the input, the model predicts the consumption in that segment, according to the class of the driver. All snippets, with their respective consumption, are returned to the API (11).

4.3.3 Integration and User Interface Components

The API is used by a mobile application that traces the route according to the user's starting point and the chosen destination. This application can use a routes API to determine the data from the stretches as accurately and with the greatest possible sampling. This API accepts route information and driver identification (9) and, as illustrated in Algorithm 1, returns the consumption of each leg, operating with a complexity

Algorithm 1: Predicting battery consumption

```
alt ← Input.start.altitude;
class ← get_class(Input.driver)
for LEG IN INPUT.PATH do
    if LEG.VARALT IS NONE then
        leg.varalt ← (leg.end.altitute − alt)/leg.distance/1000;
    cons ← model.predict({class, leg.speed, leg.varalt, Input.airconditioning,
        Input.economode});
    alt ← leg.final.altitude;
    leg.consumption ← cons ∗ distance;
```

of $O(N)$, where N is the number of legs or segments in the `Input.path`. To call this routine, you can access the following endpoint:

```
POST /predict
```

Here's an example request:

```
{
  "driver": 1,
  "airconditioning": 1,
  "economode": 0,
  "start": {
    "latitude": −21.799384,
    "longitude": −46.597957,
    "altitude": 1314
  },
  "path": [
    {
        "final": {
            "latitude":  −21.7920362,
            "longitude": −46.5937391,
            "altitude": 1292,
        },
        "speed": 7.222,
        "distance": 1.3,
        "varalt": 0,
        "consumption": 0
    }, ...
  ]
}
```

This function returns the original object while populating `varalt` and `consumption` with the altitude variation and fuel consumption of each leg, respectively. The legs of the path are provided within the `path` array, where each leg includes the final coordinate (`final`), average speed (`speed`), and length (`distance`). The driver's class is determined by consulting a database table (10) that contains driver classifications. Subsequently, the model is invoked with both the driver's classification and trip data, generating a consumption estimate for each section of the trip (11). This estimate is then relayed to the User App (12) for journey planning and recommendations regarding recharging stops.

This application is responsible for connecting the driver to the vehicle. It also has the function of tracing the travel route, starting from the selected origin and destination, and recommending stops for recharging along the way. Once the route is traced by a

Algorithm 2: Sample algorithm for application, using Google Maps API

```
steps ← gmap.directions(Input.start_coord, Input.end_coord);
lat ← −1;
long ← −1;
elevation ← −1;
for STEP IN STEPS do
    points ← step.points;
    distance ← step.distance;
    duration ← step.duration;
    elevations ←gmaps.elevation_along_path(points);
    speed ← distance/duration;
    for COORDINATE IN ELEVATIONS do
        if lat ≠ −1 or long ≠ −1 then
            step_distance ← geopy.distance((lat, long),
              (coordinate.latitude, coordinate.longitude) ).km;
            valalt = (coordinate.elevation − elevation)/step_distance/1000;
            data ← data.append({step_distance, Input.driver, speed, varalt,
              Input.airconditioning, Input.economode});
        lat ← coordinate.latitude;
        long ← coordinate.longitude;
        elevation ← coordinate.elevation;
    lat ← −1;
    long ← −1;
    elevation ← −1;
```

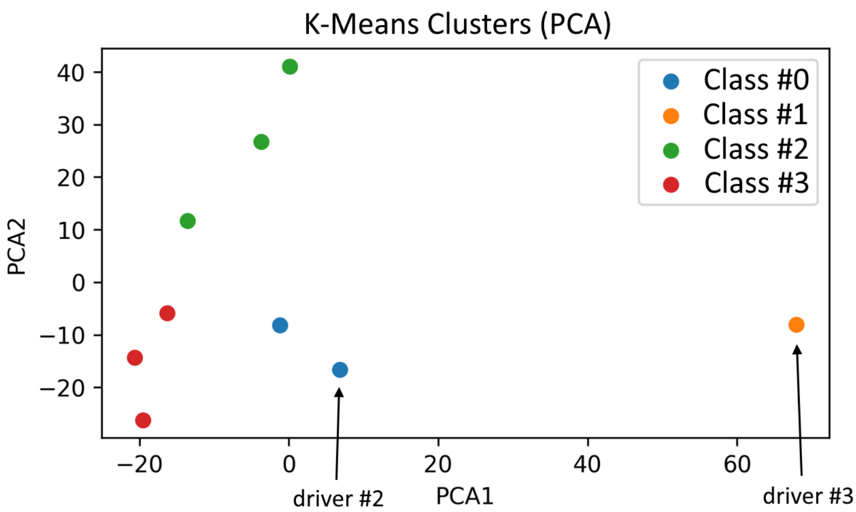
trip routing API, as detailed in Algorithm 2, the data for each segment and the driver's identification are passed to the Web API (9). The API then returns the route's expected energy expenditure (12) to support the recommendation of recharging stations. The time complexity of this data-gathering algorithm is $O(N)$, where N is the total number of coordinate points returned by the `gmaps.elevation_along_path` API. Typically, this results in a density of approximately one coordinate point per kilometer of the trip.

5 EXPERIMENTAL EVALUATION AND RESULTS

5.1 Driver Classification Analysis

After conducting the driver classification routine, we partitionated drivers into four distinct classes, as depicted in Figure 10, where drivers #2 and #3 are highlighted. The choice of $k = 4$ was guided by a convergence of empirical criteria: (i) the elbow in the within-cluster sum of squares curve occurred around $k = 4$; (ii) cluster assignments were stable under resampling; and (iii) the resulting groups were interpretable in terms of acceleration and braking behavior. Histograms of ACC, BRK, and SPD for drivers #2 and #3, in Figure 11, illustrate these behavioral differences.

Figure 10 – K-means Clustering



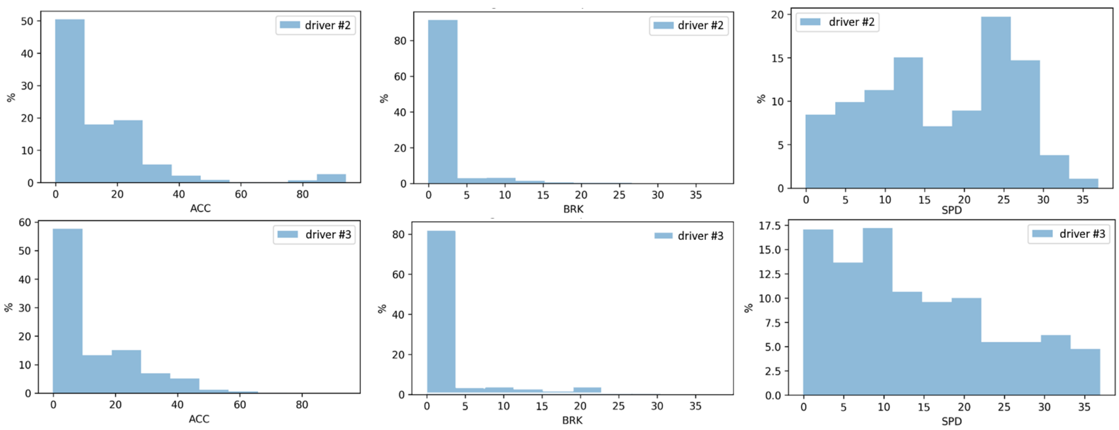
Source: The authors

The first two main components explain ~74% of the variance (~84% with three, ~93% with four). This suggests that reducing to 3–4 PCs preserves most structure for k-means, while 2 PCs are adequate for visualization.

5.2 Experimental Setup and Regression Models Evaluated

We evaluate the proposed architecture verifying ten machine learning techniques for continuous function multiple regression: *Random Forest Regressor* (RFR), *Support Vector*

Figure 11 – Histograms of accelerator pedal (ACC), brake pedal (BRK) and average speed (SPD) for drivers #2 and #3



Source: The authors

Regression (SVR), *Linear Regression* (LR), *Ridge Regression* (Ridge), *Lasso Regression* (Lasso), *Elastic Net Regression* (ENR), *K-Neighbors Regressor* (KNR), *Decision Tree Regressor* (DTR), *Ada Boost Regressor* (Ada) and *Gradient Boosting Regressor* (GBR). For all these regressors we use the default hyperparameters of the scikit-learn package, except for n_estimators in the Random Forest and ADA Boost, which we used 500, instead of 100. Also in SVR, we used the rbf kernel.

The initial evaluation of the techniques took place through 10-fold cross-validation. The indicator used in this first evaluation was the MAE, with its standard deviation. For each technique, data from four different situations were evaluated: (a) without driver information or trip context, only average speed, (b) only driver information and average speed, (c) only trip context information, and (d) with driver information and travel context. The results of the ten techniques are presented in Table 5. Subsequently, to assess sensitivity to outliers, we also computed the RMSE, again reported as mean ± standard deviation, and present these results in Table 6.

5.3 Correlation and Consumption Pattern Analysis

The analysis of the correlation matrix, shown on Figure 12, provides critical insights for any predictive modeling effort. It confirms that altitude variation (VARALT) is the most powerful linear predictor of energy consumption (CONS). Average speed (SPD) is also a significant factor.

In Figure 13, we present scatter plot of each feature against consumption per kilometer. We can observe that driver in class #1 spends more per kilometer and has the

Table 5 – Medium absolute error and standard deviation.

	(a)	(b)	(c)	(d)
RFR	1.990 ± 0.066	1.682 ± 0.061	0.287 ± 0.016	0.196 ± 0.012
SVR	1.999 ± 0.069	2.002 ± 0.060	2.005 ± 0.067	1.990 ± 0.065
LR	2.110 ± 0.069	2.106 ± 0.067	0.691 ± 0.026	0.691 ± 0.026
Ridge	2.110 ± 0.069	2.106 ± 0.067	0.998 ± 0.048	0.994 ± 0.046
Lasso	2.102 ± 0.074	2.102 ± 0.074	2.102 ± 0.074	2.102 ± 0.074
ENR	2.103 ± 0.072	2.103 ± 0.072	2.103 ± 0.072	2.103 ± 0.072
KNR	2.092 ± 0.101	1.594 ± 0.052	1.103 ± 0.062	1.144 ± 0.049
DTR	1.999 ± 0.068	1.808 ± 0.067	0.299 ± 0.025	0.198 ± 0.018
Ada	2.020 ± 0.053	1.999 ± 0.046	0.722 ± 0.029	0.641 ± 0.013
GBR	1.940 ± 0.062	1.777 ± 0.042	0.474 ± 0.019	0.360 ± 0.018

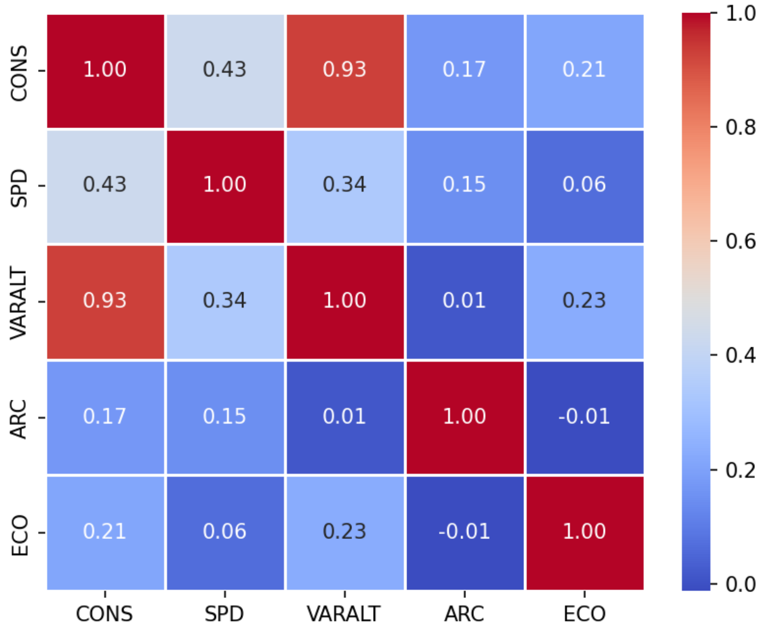
Table 6 – Root mean squared error and standard deviation.

	(a)	(b)	(c)	(d)
RFR	2.502 ± 0.078	2.318 ± 0.090	0.432 ± 0.029	0.288 ± 0.025
SVR	2.502 ± 0.077	2.486 ± 0.068	2.479 ± 0.075	2.463 ± 0.074
LR	2.578 ± 0.079	2.576 ± 0.078	0.885 ± 0.039	0.884 ± 0.039
Ridge	2.578 ± 0.079	2.576 ± 0.078	1.270 ± 0.055	1.269 ± 0.054
Lasso	2.588 ± 0.083	2.588 ± 0.083	2.588 ± 0.083	2.588 ± 0.083
ENR	2.582 ± 0.082	2.582 ± 0.082	2.582 ± 0.082	2.582 ± 0.082
KNR	2.672 ± 0.115	2.217 ± 0.079	1.606 ± 0.095	1.720 ± 0.070
DTR	2.514 ± 0.079	2.546 ± 0.098	0.556 ± 0.034	0.387 ± 0.031
Ada	2.445 ± 0.062	2.397 ± 0.044	0.872 ± 0.033	0.772 ± 0.018
GBR	2.393 ± 0.080	2.211 ± 0.065	0.623 ± 0.028	0.464 ± 0.024

higher average speed of all. We can also see that the altitude variation has the highest correlation, with mild average speed correlation. Upon further examination of the scatter plots, we have identified several key insights into the factors influencing consumption per kilometer across the four driver classes. First, we observe that altitude variation exhibits the most consistent and strongest positive linear relationship with consumption across all four classes, with Pearson correlation coefficients (r) ranging from 0.944 to 0.984. From this uniformity, we suggest that the energy required to overcome changes in elevation is a dominant and predictable factor in fuel consumption, regardless of the driver's classification.

We also note that while average speed shows a positive correlation with consumption for all classes, the relationship is considerably weaker and more dispersed (r values between 0.362 and 0.432). We find the wide spread of data points for this feature indicates that although higher speeds generally lead to higher consumption, the effect is not as direct or isolated as with altitude variation. We propose this suggests a more complex interaction where other variables, such as acceleration patterns and traffic conditions,

Figure 12 – Pearson Correlation Matrix - Consumption, Average Speed, Altitude Variation, Ar-Conditioning and Economode Usage

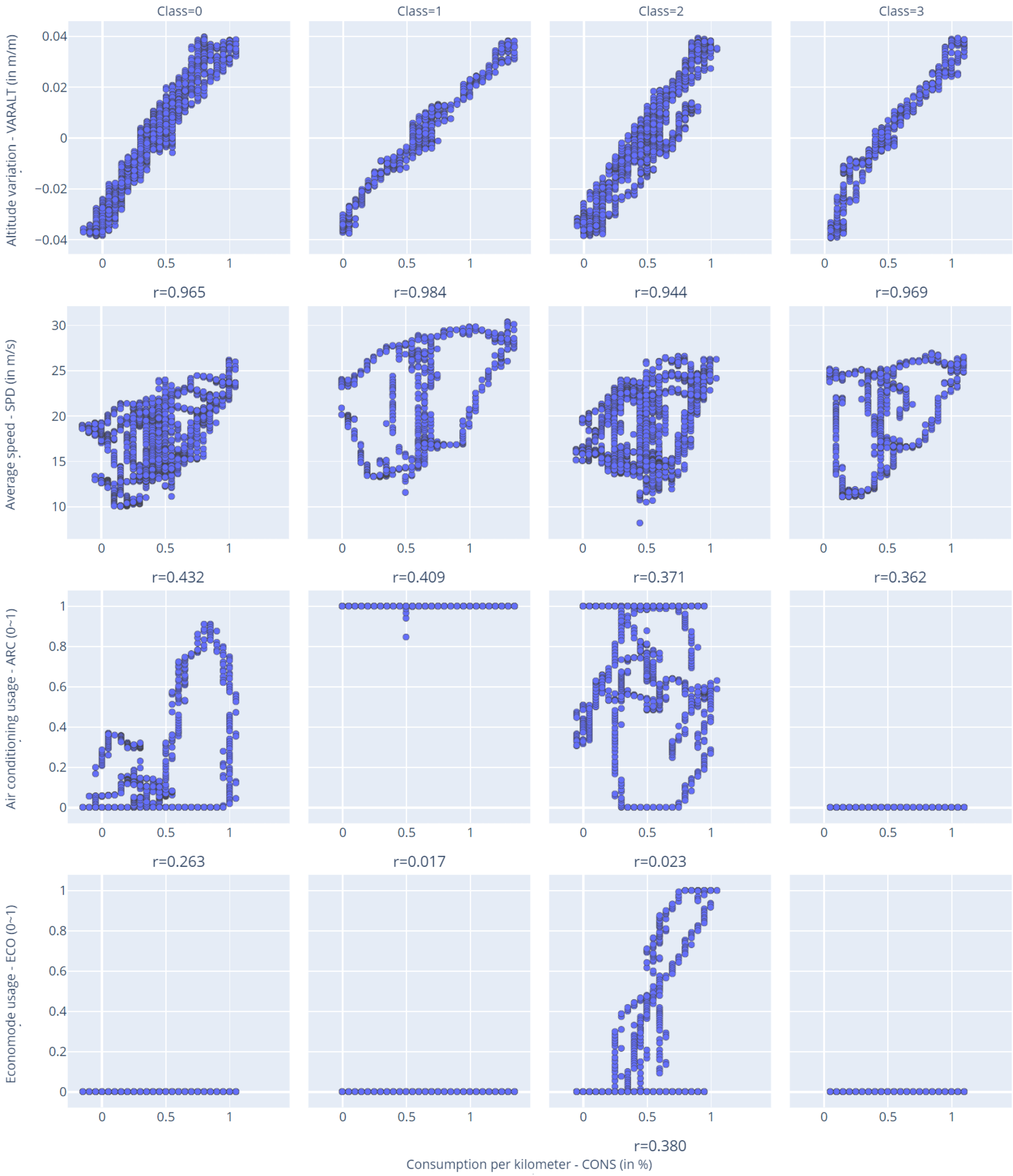


Source: The authors

heavily influence consumption at any given average speed. The binary nature of features like air conditioning (AC) and Ecomode usage provides us with a clear view of distinct behavioral patterns. Our initial analysis of Class 1 (constant AC use) and Class 3 (no AC use) is confirmed, showing how a lack of variability in a feature can lead to an uninformative correlation coefficient. For Classes 0 and 2, however, we see varied AC usage. In these cases, while a weak positive correlation exists, we observe from the plots that high consumption occurs both with and without the AC active. We therefore conclude that its use is not a primary driver of consumption variability for these drivers.

Most notably, we find the Economode plots highlight significant behavioral differences. For Classes 0, 1, and 3, we see that Economode usage is either null or consistently off, rendering a correlation analysis moot. In sharp contrast, our analysis shows that Class 2 exhibits varied use of the Economode, which correlates positively with consumption ($r = 0.380$). We suggest this counterintuitive finding indicates that drivers in this class may engage Economode during more demanding driving segments (e.g., urban traffic or high-consumption highway driving) where they perceive a greater need for fuel-saving assistance. Therefore, we posit that the mode is activated in conditions that already lead to higher consumption.

Figure 13 – Consumption per kilometer x (Altitude Variation, Average Speed, Ar-Conditioning and Economode usage)



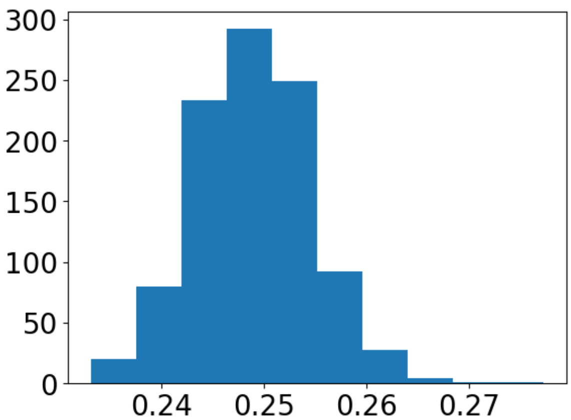
Source: The authors

5.4 Performance of Regression Models

Considering that the battery charge is obtained in steps of 0.5 pp., we evaluated that the methods that had an average error below this value are good candidates. In these tables, we observe that the RFR and DTR methods obtained better results in relation to the mean absolute error, indicating an error of 0.196 with a standard deviation of 0.012 and 0.198, with a standard deviation of 0.018 respectively, for the case where all features were used in the regression. They also got better results in the root mean squared error. When analyzing the results, we also observed that the inclusion of travel context data significantly reduces the mean absolute error, indicating that the use of these data improves the prediction made by the model. Including the identification of the conductor, a decrease in error is also noted, but in a less significant way. If only average velocity data are used, the model has similar errors across the different models, but more than doubles the 0.5 pp. accuracy of the battery charge meter.

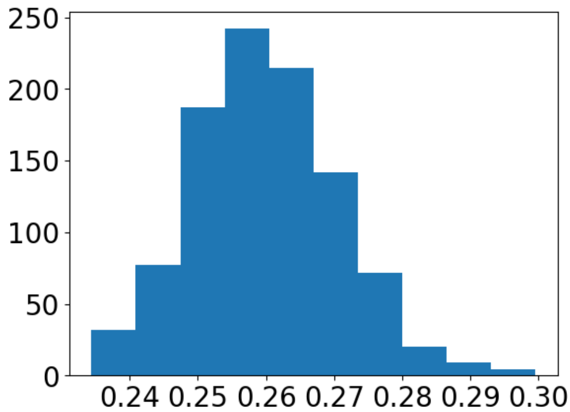
From this analysis, we applied the Bootstrap re-sampling method 1,000 times, in the dataset (d), to measure the mean error of the population. For this stage, the methods best placed in the previous evaluation (RFR and DTR) were tested. The results are shown in Figure 14 and Figure 15. Considering the 95% confidence interval, the mean absolute errors of the methods were 0.238 to 0.260 and 0.239 to 0.282, respectively. The calculated errors are still smaller than the order of magnitude of the battery charge measurement and indicate the RFR method as the best method to predict the consumption.

Figure 14 – MAE Random Forest



Source: The authors

Figure 15 – MAE Decision Tree



Source: The authors

5.5 Model Validation with Real-World and API Data

In the final phase of our study, we conducted a comprehensive comparison between the actual energy consumption of each trip and the predicted consumption generated by the regressor. The results of this comparison are presented in detail in Table 7.

Table 7 – Comparision between real and predicted consumption

Driver	Real discharge	Predicted discharge	Difference
#1	36.0	38.1	+2.1
#2	37.5	36.5	-1.0
#3	46.0	44.4	-1.6
#4	35.5	36.3	+0.8
#5	24.0	25.4	+1.4
#6	33.0	33.8	+0.8
#7	38.0	39.8	+1.8
#8	35.5	35.8	+0.3
#9	12.5	10.3	-2.2

In this extended study, we aimed to not only validate the performance of our machine learning model but also to mitigate overfitting by incorporating external data for predictions. Specifically, we utilized route calculations and altitude data from Google Maps API to predict battery consumption. This approach helps ensure that the model's predictions are generalizable and not overly tailored to the training data. This phase aimed to assess the accuracy of the predicted consumption versus the real consumption recorded for seven drivers. Two drivers were not considered as their telemetry data was not complete. Each driver's route was mapped and the consumption was calculated based on the distance, altitude and estimated travel time provided by the Google Maps API. The differences between the actual consumption and each prediction method were analyzed to determine the accuracy and reliability of the predictions, as shown in Table 8.

Table 8 – Comparison between real and predicted consumption with Google API route information

Driver	Real discharge	Predicted discharge	Difference
#1	36.0	33.7	-2.3
#2	37.5	33.5	-4.0
#3	46.0	37.9	-8.1
#4	35.5	33.4	-2.1
#6	33.0	30.2	-2.8
#7	38.0	35.6	-2.4
#8	35.5	31.8	-3.7

5.6 Analysis of the Route Correction Factor

To further analyze the discrepancies and understand the underlying causes, we divided the route into 18 stretches and analyzed the time each stretch took to accomplish. We compared these times with the actual time taken by each driver. The analysis is summarized in Table 9, showing the relation between the predicted and actual times for each leg.

Table 9 – Comparison of leg completion times between Google predicted and driver's actual times

Leg	Google	#1	#2	#3	#4	#6	#7	#8
1	3.00	4.07	2.67	4.42	5.33	3.92	8.40	4.60
2	3.00	5.05	3.17	4.33	3.28	3.35	4.30	3.38
3	4.00	2.68	2.78	2.48	3.37	3.12	3.48	2.73
4	5.00	3.92	3.73	3.25	3.82	4.35	4.07	4.27
5	6.00	4.68	4.37	3.82	4.80	5.27	4.55	5.35
6	10.00	9.05	7.60	7.70	11.08	10.33	11.25	9.48
7	4.00	3.23	4.73	4.20	3.45	7.02	3.30	4.03
8	5.00	5.32	6.83	5.33	4.40	5.60	4.67	5.33
9	4.00	4.27	4.18	3.97	3.80	4.52	3.70	4.02
10	5.00	3.93	3.85	3.22	3.67	4.63	3.25	3.67
11	4.00	4.12	3.27	3.47	4.02	3.53	4.60	3.93
12	3.00	2.78	2.38	2.70	2.68	3.03	2.75	2.57
13	10.00	8.87	7.63	6.68	7.77	9.15	8.27	7.95
14	6.00	4.40	4.25	3.65	4.73	4.67	4.70	4.82
15	5.00	3.77	3.65	3.15	3.90	4.23	3.98	4.68
16	4.00	3.75	2.40	3.85	3.95	3.85	2.98	3.62
17	2.00	1.97	1.15	1.37	1.75	1.55	2.25	1.68
18	3.00	2.68	2.52	3.43	3.60	3.28	3.83	3.13
Total	86.00	78.53	71.17	71.02	79.40	85.40	84.33	79.25

We found that the variations in time taken for each stretch by different drivers significantly influenced the prediction accuracy. These time differences affect the average speed, which is a critical factor in predicting battery consumption accurately. To address this, we adjusted our prediction model to account for the actual times taken by drivers on each stretch, using a factor $k_{i,j}$ for each leg i and driver j. By incorporating these adjustments, we observed a significant improvement in the prediction accuracy. This suggests that considering the individual driving patterns and the specific time taken on each segment can lead to more precise predictions of battery consumption, as we can see in Table 10.

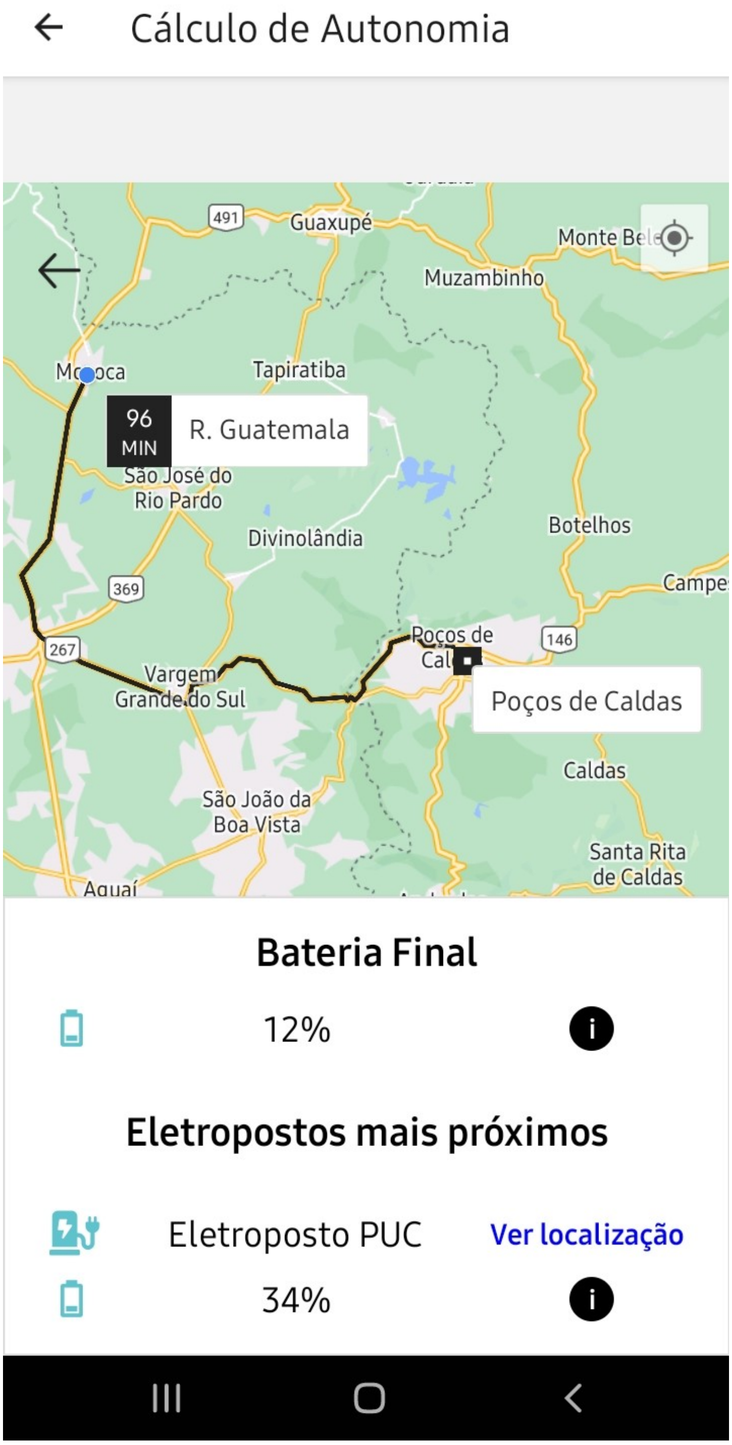
Table 10 – Comparison between real and predicted consumption, using factor $k_{i,j}$ along the route

Driver	Real discharge	Predicted discharge	Difference
#1	36.0	36.9	+0.9
#2	37.5	37.1	-0.4
#3	46.0	44.4	-1.6
#4	35.5	35.8	+0.3
#6	33.0	31.2	-1.8
#7	38.0	38.9	+0.9
#8	35.5	33.8	-1.7

5.7 User Application Demonstration

Furthermore, to show the versatility of our API for integration with third-party applications, our research involved the utilization of a User App prototype called MOV+. This innovative tool empowered us to offer valuable recharging station recommendations proactively, well before the battery charge level reaches a critical 34%, as depicted in Figure 16 (Note: The figure is in Portuguese).

Figure 16 – Simulated travel route - 104 km - User App MOV+, displaying autonomy prediction (Cálculo de Autonomia), final battery charge (Bateria Final) until target, 12%, and the nearest charging station (i.e., Eletroposto PUC), at 34%



Source: The authors

6 CONCLUSION

With the increasing adoption of electric vehicles (EVs) worldwide, challenges related to limited driving range and charging infrastructure are becoming more pronounced. Accurate range prediction and strategic charging point recommendations are essential for enhancing the EV user experience.

Our proposed architecture addresses these challenges by leveraging telemetry data to classify drivers based on their driving behavior and analyzing trip segment data to predict energy consumption and range, to help applications to suggest optimal charging stops along the route. Machine learning techniques are employed for driver classification and consumption prediction models.

Our analysis reveals that the Random Forest regression method can effectively predict battery consumption for a given trip segment when driver and contextual data are available. Routing applications provide crucial information such as altitude variations and average speed, while additional data like driver behavior, air conditioning usage, and economy mode settings are inputted via an app. The primary objective of this study —to validate the proposed architecture's feasibility — was achieved.

Our findings indicate that prediction accuracy is heavily dependent on data precision. For instance, with battery charge variations as small as 0.5 percentage points (pp), a mean absolute error of 0.249 pp is deemed acceptable. However, there is potential for further accuracy improvements. A significant discrepancy was noted between estimated and actual travel times, impacting average speed and subsequent consumption predictions.

In conclusion, our research underscores the importance of considering individual driving styles and contextual factors in developing reliable battery consumption predictive models. We identified the necessity of applying a correction factor to the average speed of each significant trip segment. This factor, termed $k_{i,j}$, represents the ratio of actual average speed to that predicted by routing applications.

Additionally, expanding the dataset to include more instances of economy mode usage — currently less than 10% of the total dataset — could enhance prediction precision.

6.1 Future Work

Future work should explore whether K-Means is the optimal algorithm for classifying driving modes and determine the ideal number of driver classes, given the limited driver data in this study. Additional features —such as acceleration patterns, steering behavior, and braking dynamics— should also be examined to improve the accuracy of mode classification.

Further investigation is needed to assess the impact of dataset time intervals on regression model training, as the collected data spans a limited period. Examining additional contextual factors such as temperature, travel time, and geographic coordinates may also improve consumption forecasts.

Finally, further studies should explore the relationship between the $k_{i,j}$ factor and driver behavior. Methods to calculate this factor by comparing the collected driving data with official speed limit information for each road segment might be an improvement.

6.2 Contributions

The main contributions of this dissertation are centered on the conception, development, and validation of a complete architecture for predicting the autonomy of electric vehicles. The primary contributions include:

- The design of a modular system architecture that uniquely integrates a driver classification module, based on driving style (drivability) , with a consumption prediction model that uses trip context data, such as altitude variation and average speed.
- The creation and public dissemination of a detailed telemetry dataset collected from a real-world electric vehicle, addressing a significant gap and providing a valuable resource for the research community.
- A comprehensive evaluation of ten different machine learning regression algorithms, identifying the Random Forest Regressor as the most effective method for predicting battery consumption with high precision under the studied conditions.

The research conducted during the development of this dissertation resulted in the following publications: Reis, Teixeira e Marques-Neto (2023), presented in Portuguese, introduced the foundational concepts of the ArchDriva architecture. It presented the initial proposal that integrates a driver classification algorithm with a regression model to determine energy consumption, forming the basis for the more extensive validation and in-depth results detailed in this dissertation. And, Reis et al. (2024), published in an international journal, details the proposed software architecture for trip planning and routing of electric vehicles. It provides an in-depth evaluation of regression methods for consumption prediction and validates the architecture by comparing real trips with simulations, using the dataset generated in this research. This work represents a consolidation and validation of the main technical results of the dissertation.

Acknowledgements

This work was developed with the support of FAPEMIG (PPM-XII), DME Distribuição (DMED) and the Agência Nacional de Energia Elétrica (ANEEL).

Availability of data and materials

The datasets generated and/or analysed during the current study are available in https://github.com/mreis76/electric-vehicle-telemetry-dataset

REFERENCES

ADEDEJI, B. P. Electric vehicles survey and a multifunctional artificial neural network for predicting energy consumption in all-electric vehicles. Results in Engineering, v. 19, p. 101283, 2023. Disponível em: <https://doi.org/10.1016/j.rineng.2023.101283>.

ADEDEJI, B. P. Electric vehicles survey and a multifunctional artificial neural network for predicting energy consumption in all-electric vehicles. Results in Engineering, v. 19, p. 101283, 2023. ISSN 2590-1230. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2590123023004103>.

ADEDEJI, B. P.; KABIR, G. A feedforward deep neural network for predicting the state-of-charge of lithium-ion battery in electric vehicles. Decision Analytics Journal, v. 8, p. 100255, 2023. ISSN 2772-6622. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2772662223000954>.

ADEDEJI, B. P.; KABIR, G. A feedforward deep neural network for predicting the state-of-charge of lithium-ion battery in electric vehicles. Decision Analytics Journal, v. 8, p. 100255, 2023. Disponível em: <https://doi.org/10.1016/j.dajour.2023.100255>.

AKL, M. M.; AHMED, A. A.; RASHAD, E. E. M. A wide component sizing and performance assessment of electric drivetrains for electric vehicles. In: IEEE. 2019 21st International Middle East Power Systems Conference (MEPCON). [S.l.], 2019. p. 834–839.

ALANAZI, F. Electric vehicles: Benefits, challenges, and potential solutions for widespread adaptation. Applied Sciences, v. 13, n. 10, 2023. ISSN 2076-3417. Disponível em: <https://www.mdpi.com/2076-3417/13/10/6016>.

ALI, I.; NAUSHAD, M. A study to investigate what tempts consumers to adopt electric vehicles. World Electric Vehicle Journal, v. 13, n. 2, p. 26, 01 2022.

ALPAYDIN, E. Introduction to machine learning. [S.l.]: MIT press, 2020.

ANATEL. Anatel - Infraestrutura. 2022. `https://informacoes.anatel.gov.br/paineis/infraestrutura/`, Accessed 7 January 2022.

ARABIE, P.; HUBERT, L.; SOETE, G. D. Clustering and classification. [S.l.]: World Scientific, 1996.

BAILEY, C. et al. Electric vehicle autonomy: Realtime dynamic route planning and range estimation software. In: 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC). [S.l.: s.n.], 2022. p. 2696–2701.

BOLOVINOU, A. et al. Online prediction of an electric vehicle remaining range based on regression analysis. In: IEEE. 2014 IEEE International Electric Vehicle Conference (IEVC). [S.l.], 2014. p. 1–8.

BOOS, D. D. Introduction to the Bootstrap World. Statistical Science, Institute of Mathematical Statistics, v. 18, n. 2, p. 168 – 174, 2003. Disponível em: <https://doi.org/10.1214/ss/1063994971>.

BRATAAS, G.; HUGHES, P. Exploring architectural scalability. SIGSOFT Softw. Eng. Notes, Association for Computing Machinery, New York, NY, USA, v. 29, n. 1, p. 125–129, jan. 2004. ISSN 0163-5948. Disponível em: <https://doi-org.ez93.periodicos.capes.gov.br/10.1145/974043.974064>.

BUSTOS, J. E. G. et al. A novel data-driven framework for driving range prognostics in electric vehicles. Engineering Applications of Artificial Intelligence, v. 142, p. 109925, 2025. ISSN 0952-1976. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0952197624020840>.

CHAI, T.; DRAXLER, R. R. Root mean square error (rmse) or mean absolute error (mae). Geoscientific Model Development Discussions, v. 7, n. 1, p. 1525–1534, 2014.

CHANDRAN, V. et al. State of charge estimation of lithium-ion battery for electric vehicles using machine learning algorithms. World Electric Vehicle Journal, v. 12, n. 1, 2021. ISSN 2032-6653. Disponível em: <https://www.mdpi.com/2032-6653/12/1/38>.

Chen, M.-S. et al. Driving behaviors analysis based on feature selection and statistical approach: a preliminary study. The Journal of Supercomputing, Springer, v. 75, n. 4, p. 2007–2026, 2019.

Chen, S. et al. On the relationship between energy consumption and driving behavior of electric vehicles based on statistical features. In: 2019 Chinese Control Conference (CCC). [S.l.: s.n.], 2019. p. 3782–3787.

Chen, W.; Lin, Y.; Chen, W. Comparisons of machine learning algorithms for driving behavior recognition using in-vehicle can bus data. In: 2019 Joint 8th International Conference on Informatics, Electronics Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision Pattern Recognition (icIVPR). [S.l.: s.n.], 2019. p. 268–273.

Delnevo, G. et al. On combining big data and machine learning to support eco-driving behaviours. Journal of Big Data, SpringerOpen, v. 6, n. 1, p. 64, 2019.

EBERLY, L. E. Multiple linear regression. Topics in Biostatistics, Springer, p. 165–187, 2007.

FUGIGLANDO, U. et al. Driving behavior analysis through can bus data in an uncontrolled environment. IEEE Transactions on Intelligent Transportation Systems, IEEE, v. 20, n. 2, p. 737–748, 2018.

Fugiglando, U. et al. Driving behavior analysis through can bus data in an uncontrolled environment. IEEE Transactions on Intelligent Transportation Systems, v. 20, n. 2, p. 737–748, 2019.

GÖNÜL, Ö.; DUMAN, A. C.; GÜLER, Ö. Electric vehicles and charging infrastructure in turkey: An overview. Renewable and Sustainable Energy Reviews, Elsevier, v. 143, p. 110913, 2021.

GOUJON, A.; HAGHIGHI, H.; HAMRI, M. E. A. Impact of eco-driving on energy consumption of an electric vehicle. In: 2022 7th International Conference on Environment Friendly Energies and Applications (EFEA). [S.l.: s.n.], 2022. p. 1–6.

HALONEN, T.; ROMERO, J.; MELERO, J. GSM, GPRS and EDGE performance: evolution towards 3G/UMTS. [S.l.]: John Wiley & Sons, 2004.

Huang, Y. et al. Save or waste: Real data based energy-efficient driving. IEEE Access, v. 8, p. 133936–133950, 2020.

HUCHO, W. Commercial vehicles. Aerodynamics of Road Vehicles, fourth ed., SAE International, Warrendale, Pa., United States, p. 131–237, 1998.

International Energy Agency. Global EV Outlook 2025. 2025. Available in `https://www.iea.org/reports/global-ev-outlook-2025`. Accessed on: 18 July 2025.

ISO Central Secretary. Road vehicles -— Controller area network (CAN) -— Part 1: Data link layer and physical signalling. Geneva, CH, 2015. Disponível em: <https://www.iso.org/standard/63648.html>.

Jia, S. et al. Long short-term memory and convolutional neural network for abnormal driving behaviour recognition. IET Intelligent Transport Systems, v. 14, n. 5, p. 306–312, 2020.

JIANG, Y. et al. Remaining driving range prediction of electric vehicles based on personalized driving behavior in complex traffic scenarios. IEEE Open Journal of Vehicular Technology, p. 1–15, 2025.

KOCAARSLAN, I. et al. High-fidelity electric vehicle energy consumption modelling and investigation of factors in driving on energy consumption. In: 2022 4th Global Power, Energy and Communication Conference (GPECOM). [S.l.: s.n.], 2022. p. 227–231.

Krishnamurthy, V. et al. Classification of driving behavior events utilizing kinematic classification and machine learning for down sampled time series data. In: 2019 IEEE International Conference on Big Data (Big Data). [S.l.: s.n.], 2019. p. 3789–3796.

LEBROUHI, B. et al. Key challenges for a large-scale development of battery electric vehicles: A comprehensive review. Journal of Energy Storage, Elsevier, v. 44, p. 103273, 2021.

Li, H.; Kang, H. Research on user behavior prediction and profiling method based on trajectory information. Automatic Control and Computer Sciences, Springer, v. 54, n. 5, p. 456–465, 2020.

LIU, J.; SHI, P.; FAN, J. Electric vehicle energy consumption model based on working condition recognition and k-means cluster analysis. In: 2022 INTERNATIONAL CONFERENCE ON MANUFACTURING, INDUSTRIAL AUTOMATION AND ELECTRONICS (ICMIAE). [S.l.: s.n.], 2022. p. 130–134.

Liu, S.; Wang, N.; Wang, X. Multi-factor influenced integrated driving behavior clustering analysis evaluation model. In: 2019 INTERNATIONAL CONFERENCE ON COMPUTER, INFORMATION AND TELECOMMUNICATION SYSTEMS (CITS). [S.l.: s.n.], 2019. p. 1–5.

MASHKOV, V.; KAROVA, M. Estimation of power consumption for an electric vehicle through variable parameter modelling. In: 2023 INTERNATIONAL CONFERENCE AUTOMATICS AND INFORMATICS (ICAI). [S.l.: s.n.], 2023. p. 119–124.

MASHKOV, V.; KAROVA, M. Estimation of power consumption for an electric vehicle through variable parameter modelling. In: 2023 INTERNATIONAL CONFERENCE AUTOMATICS AND INFORMATICS (ICAI). [s.n.], 2023. p. 119–124. Disponível em: <https://doi.org/10.1109/ICAI58806.2023.10339079>.

MCKINSEY. O FUTURO DA MOBILIDADE NO BRASIL: UMA ROTA PARA ELETRIFICAÇÃO. 2023. `https://www.mckinsey.com.br/our-insights/all-insights/o-futuro-da-mobilidade-no-brasil`, Accessed 27 January 2023.

MUSTAFFA, Z.; SULAIMAN, M. H.; ISUWA, J. State of charge estimation of lithium-ion batteries in an electric vehicle using hybrid metaheuristic - deep neural networks models. ENERGY STORAGE AND SAVING, 2025. ISSN 2772-6835. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2772683525000068>.

NOUSSAN, M. PUBLIC EV CHARGING IN ITALY: POLICIES, BUSINESS MODELS AND FUTURE SCENARIOS. [S.l.]: MDPI, 2020.

PAUL, D. et al. Predicting energy consumption of battery-operated electric vehicles: A comparative performance assessment. In: 2023 IEEE INTERNATIONAL CONFERENCE ON INDUSTRIAL ENGINEERING AND ENGINEERING MANAGEMENT (IEEM). [s.n.], 2023. p. 1032–1036. Disponível em: <https://doi.org/10.1109/IEEM58616.2023.10406558>.

PAUL, D. et al. Predicting energy consumption of battery-operated electric vehicles: A comparative performance assessment. In: 2023 IEEE INTERNATIONAL CONFERENCE ON INDUSTRIAL ENGINEERING AND ENGINEERING MANAGEMENT (IEEM). [S.l.: s.n.], 2023. p. 1032–1036.

Ping, P. et al. Impact of driver behavior on fuel consumption: Classification, evaluation and prediction using machine learning. IEEE ACCESS, v. 7, p. 78515–78532, 2019.

PRASAD, S. L.; GUDIPALLI, A. An effective range estimation and state-of-charge to mitigate range anxiety in electric vehicles. HELIYON, v. 11, n. 1, p. e41494, 2025. Disponível em: <https://doi.org/10.1016/j.heliyon.2024.e41494>.

R., D.; VENKATESHKUMAR, M.; KUPPUSAMY, C. L. Driving range estimation using amrit for prognostics model of electric vehicles. In: 2023 INNOVATIONS IN POWER AND ADVANCED COMPUTING TECHNOLOGIES (I-PACT). [s.n.], 2023. p. 1–6. Disponível em: <https://doi.org/10.1109/i-PACT58649.2023.10434455>.

REIS, M. dos et al. Using electric vehicle driver's driving mode for trip planning and routing. Journal of Internet Services and Applications, v. 15, n. 1, p. 410–423, Sep. 2024. Disponível em: <https://journals-sol.sbc.org.br/index.php/jisa/article/view/3805>.

REIS, M. dos; TEIXEIRA, F.; MARQUES-NETO, H. Utilizando o modo de dirigir do motorista de veículo elétrico para o planejamento e roteirização de viagem. In: Anais do VII Workshop de Computação Urbana. Porto Alegre, RS, Brasil: SBC, 2023. p. 15–28. ISSN 2595-2706. Disponível em: <https://sol.sbc.org.br/index.php/courb/article/view/24568>.

RODRIGUEZ, M. Z. et al. Clustering algorithms: A comparative approach. PloS one, Public Library of Science San Francisco, CA USA, v. 14, n. 1, p. e0210236, 2019.

Sama, K. et al. Extracting human-like driving behaviors from expert driver data using deep learning. IEEE Transactions on Vehicular Technology, v. 69, n. 9, p. 9315–9329, 2020.

Siami, M.; Naderpour, M.; Lu, J. A mobile telematics pattern recognition framework for driving behavior extraction. IEEE Transactions on Intelligent Transportation Systems, p. 1–14, 2020.

SOUSA, Á. Coeficiente de correlação de pearson e coeficiente de correlação de spearman: o que medem e em que situações devem ser utilizados? Correio dos Açores, Gráfica Açoreana, Lda, p. 19–19, 2019.

Stanton, N. A.; Allison, C. K. Driving towards a greener future: an application of cognitive work analysis to promote fuel-efficient driving. Cognition, Technology & Work, Springer, v. 22, n. 1, p. 125–142, 2020.

THORGEIRSSON, A. T. et al. Probabilistic prediction of energy demand and driving range for electric vehicles with federated learning. IEEE Open Journal of Vehicular Technology, IEEE, v. 2, p. 151–161, 2021.

TIWARY, A.; GARG, S.; MISHRA, S. Impact of driving behaviour on energy consumption of electric vehicle. In: 2022 22nd National Power Systems Conference (NPSC). [S.l.: s.n.], 2022. p. 872–877.

TRAN, Q. T. et al. An improved method to estimate energy consumption for light duty electric vehicle based on road characteristics and ambient temperature. In: 2023 Asia Meeting on Environment and Electrical Engineering (EEE-AM). [S.l.: s.n.], 2023. p. 01–04.

VARGAS, J. E. V. et al. The new neighbor across the street: An outlook for battery electric vehicles adoption in brazil. World Electric Vehicle Journal, v. 11, n. 3, 2020. ISSN 2032-6653. Disponível em: <https://www.mdpi.com/2032-6653/11/3/60>.

WANG, L. et al. Enhancing electric vehicle energy consumption prediction: Integrating elevation into machine learning model. In: 2024 IEEE Intelligent Vehicles Symposium (IV). [S.l.: s.n.], 2024. p. 2936–2941.

Warren, J.; Lipkowitz, J.; Sokolov, V. Clusters of driving behavior from observational smartphone data. IEEE Intelligent Transportation Systems Magazine, v. 11, n. 3, p. 171–180, 2019.

WEI, H. et al. Online estimation of driving range for battery electric vehicles based on soc-segmented actual driving cycle. Journal of Energy Storage, v. 49, p. 104091, 2022. ISSN 2352-152X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352152X22001281>.