



**ESTIMATION DU TEMPS DE CONDUITE D'UN VÉHICULE ÉLECTRIQUE ENTRE DEUX ETATS
DE CHARGE À L'AIDE DES ALGORITHMES D'APPRENTISSAGE AUTOMATIQUE**

Par Chourik Fousseni

**Mémoire présenté à l'Université du Québec à Chicoutimi en vue de l'obtention du grade de
Maîtrise ès sciences appliquées (M. Sc. A.) en ingénierie**

Jury:

Martin J.-D. Otis, ing. M.Sc.A. Ph.D, Professeur, Université du Québec à Chicoutimi, Directeur de recherche

Khaled Ziane, Ph.D, Directeur, Centre de Recherche et d'Innovation en Intelligence Énergétique, Co-directeur de recherche

Maxime Berger, ing. Ph.D., Professeur, Université de sherbrooke, examinateur externe

Damien Brun, Professeur, Université du Québec à Chicoutimi, examinateur externe

QUÉBEC, CANADA

© Chourik Fousseni, 2026

RÉSUMÉ

Estimer avec précision le temps de conduite restant (*Remaining Driving Time*, RDT) de la batterie d'un véhicule électrique (VE) est essentiel pour optimiser la gestion de l'énergie et améliorer l'expérience utilisateur. Cependant, les méthodes d'estimation traditionnelles ne prennent pas suffisamment en compte l'influence de la température, des styles de conduite et du temps de conduite du véhicule, ce qui entraîne des prédictions moins précises et une gestion de l'autonomie sous-optimale.

Pour remédier à ces limites, cette étude propose une méthode d'estimation de la durée de conduite restante en intégrant la température et les styles de conduite, ce qui affine les prédictions et améliore la fiabilité du modèle. De plus, des données issues de l'Alliance Nationale des Données Massives pour les Véhicules à Nouvelles Énergies (NDANEV) ont été utilisées pour développer un modèle prédictif basé sur des algorithmes d'apprentissage automatique (*Machine Learning*, ML).

Les différents modèles ML comparés dans cette étude sont : LSTM, Random Forest, Prophet, LightGBM et XGBoost. Les performances des modèles ont été évaluées à l'aide de l'erreur absolue moyenne (*Mean Absolute Error*, MAE), de l'erreur quadratique moyenne (*Root Mean Square Error*, RMSE), du coefficient de détermination (R^2) et du temps d'exécution de la prédiction afin d'évaluer la précision des estimations.

Les résultats montrent que les valeurs de R^2 pour Prophet, Random Forest, LSTM, XGBoost et LightGBM sont respectivement de 0,91 ; 0,94 ; 0,95 ; 0,94 et 0,94. Cela suggère que LSTM surpasse les autres modèles en termes de précision, offrant l'estimation du temps de conduite restant. En outre, le résultat confirme que la prise en compte des caractéristiques de conduite et de la température ambiante améliore la fiabilité et la robustesse des estimations. Ces avancées contribuent à une gestion de l'énergie plus efficace et à des stratégies de recharge optimisées.

Mots-clés: Véhicule électrique (VE) ; état de charge (SoC) ; conditions et caractéristiques de conduite ; batterie lithium-ion (LIB) ; apprentissage automatique (ML) ; Random Forest (RF) ; mémoire à long court terme (LSTM).

TABLE DES MATIÈRES

RÉSUMÉ	ii
TABLE DES MATIÈRES.....	iii
LISTES DES TABLEAUX	v
LISTE DES FIGURES	vi
LISTES DES SIGLES.....	vii
DÉDICACE	viii
REMERCIEMENTS	ix
PLAN DU MÉMOIRE.....	x
CONTRIBUTIONS DES AUTEURS	xi
STATUT DU MANUSCRIT	xii
AVANT-PROPOS.....	xiii
CHAPITRE 1	1
1. INTRODUCTION	1
1.1. Contexte	2
1.2. Problématique	3
1.3. Objectifs	3
1.4. Contribution.....	4
1.5. Organisation des chapitres	6
CHAPITRE 2	7
2. REVUE DE LA LITTÉRATURE	7
2.1. Définitions et contextualisation des concepts	10
2.1.1. SoC	11
2.1.2. SoH	11
2.1.3. Style de conduite.....	11
2.1.4. Conditions de conduite.....	11
2.2. Analyse bibliométrique	12
2.3. Analyse des données utilisées dans les études précédentes	14
RÉSUMÉ ÉTENDU	16
CHAPITRE 3	17
MÉTHODOLOGIE	17
INTRODUCTION.....	18
3. METHODOLOGY	22
3.1. Data collection	24
3.2. Data analysis	26
3.3. K- means clustering	27

3.4.	RDT predictions models with justifications	29
3.4.1.	XGBoost	30
3.4.2.	LightGBM	31
3.4.3.	RF.....	31
3.4.4.	LSTM.....	32
3.4.5.	Prophet.....	32
3.5.	Cross-validation	33
3.6.	Hyperparameter selection.....	34
3.7.	Performances metrics	35
3.7.1.	Mean Absolute Error (MAE)	35
3.7.2.	Root Mean Squared Error (RMSE)	36
3.7.3.	Coefficient of Determination (R^2)	36
3.8.	Statistical tests	37
CHAPITRE 4		39
4.	RESULTS AND DISCUSSIONS.....	39
4.1.	Cross-validation	39
4.2.	Features selection.....	45
4.3.	Friedman test	46
4.4.	Nemenyi test	47
4.4.1.	Nemenyi test for R^2	47
4.4.2.	Nemenyi test for RMSE	48
4.4.3.	Nemenyi test for MAE	49
4.5.	Prediction of driving time for end SoC=0	50
CONCLUSION		52
CHAPITRE 5		55
5.	CONCLUSION GÉNÉRALE.....	55
LISTE DE RÉFÉRENCES.....		57
CERTIFICATION ÉTHIQUE		61

LISTES DES TABLEAUX

Table 1: Comparaison des travaux de recherche anterieurs selon leurs contriubutions	5
Table 2: Comparison of our results with existing literature works	22
Table 3: Comparison of existing techniques: advantages and limitations.....	22
Table 4: Summary of Vehicle Characteristics and Collection Period	24
Table 5: Description of dataset features.....	25
Table 6: Symbols and definitions.....	26
Table 7: Parameters values used in the machine learning models.....	35
Table 8: Results of the cross-validation experiment.....	43
Table 9: Friedman test results	46
Table 10: Performance metrics	51

LISTE DES FIGURES

Figure 1: Interaction entre les variables influençant la performance de la batterie d'un VE	4
Figure 2: Taxonomie des approches et algorithmes	12
Figure 3: Analyse VOS Viewer de publication mondiale	13
Figure 4: Flowchart for methodology used	23
Figure 5: Number of data points per group	28
Figure 6: Elbow method for determining the optimal number of clusters.....	28
Figure 7: The five cluster group	29
Figure 8: Data splitting into training and test sets for cross-validation.....	34
Figure 9: Error dynamics during real-time evaluation	40
Figure 10: Predicted vs actual values for training and test sets	41
Figure 11: Predicted time vs distance for full battery discharge.....	41
Figure 12: Error dynamics of each model during real-time evaluation	42
Figure 13: Cross-validation performance on each fold.	44
Figure 14: Cross-validation real-time tracking of prediction error	45
Figure 15: SHAP Summary Plot of Feature Impact on Model Output	46
Figure 16: Nemenyi test for R square	48
Figure 17: Nemenyi test for RMSE	49
Figure 18: Nemenyi test for MAE	50

LISTES DES SIGLES

Sigles	Significations	
	Français	Anglais
LAR.i	Laboratoire d'Automatique et de Robotique interactive	<i>Automatic and Interactive Robotics Laboratory</i>
CR2ie	Centre de Recherche et d'innovation en Intelligence Énergétique	<i>Center for Research and Innovation in Energy Intelligence</i>
NDANEV	Alliance nationale des mégadonnées pour les véhicules à énergie nouvelle	<i>National Big Data Alliance for New Energy Vehicles</i>
SoH	État de Santé	<i>State of Health</i>
SoC	État de Charge	<i>State of Charge</i>
VE	Véhicule Électrique	<i>Electric Vehicle</i>
BVE	Batterie des Véhicules Électriques	<i>Electric Vehicle Battery</i>
RDT	Durée de conduite restante	<i>Remaining Driving Time</i>
MLP	Perceptron Multicouche	<i>MultiLayer Perceptron</i>
LSTM	Mémoire à Long Terme	<i>Long Short-Term Memory</i>
TCN	Réseaux neuronaux convolutifs temporels	<i>Temporal Convolutional Neural Networks</i>
BMS	Système de Gestion de la Batterie	<i>Battery Managment System</i>
LIB	Batterie au Lithium-ion	<i>Lithium-Ion Battery</i>
LightGBM	Machine de Gradient Boosting Léger	<i>Light Gradient Boosting Machine</i>
XGBoost	Gradient Boosting Extrême	<i>Extreme Gradient Boosting</i>
RDR	Autonomie restante	<i>Remaining Driving Range</i>
RDE	Énergie restante à décharger	<i>Remaining Driving Energy</i>
ECR	Taux de consommation d'énergie	<i>Energy Consumption Rate</i>

DÉDICACE

À la douce mémoire de Micheline Massan Okamba, ton absence physique n'a jamais éteint la lumière de ta présence. Ton amour, ta sagesse et ta force tranquille résonnent encore en moi à chaque étape de ce voyage. Ce mémoire est un humble hommage à ton héritage, que je porte avec respect, gratitude et tendresse.

À mes parents, piliers de ma vie, pour leur amour sans condition et leur foi constante en mes capacités.

À mes frères et sœurs, pour leur soutien indéfectible, leur affection et leur présence rassurante tout au long de ce parcours.

À mes enseignants et mentors, dont les conseils avisés et l'exigence bienveillante ont nourri ma progression.

À mes amis, pour leur patience, leur écoute et leur encouragement tout au long de cette aventure.

Et à toutes celles et ceux qui ont cru en moi et m'ont inspiré à viser l'excellence, merci d'avoir accompagné ce chemin avec bienveillance.

REMERCIEMENTS

Je souhaite tout d'abord exprimer ma profonde gratitude au Professeur Martin Otis pour avoir accepté d'encadrer ce mémoire. Son accompagnement rigoureux, tant sur le plan scientifique que personnel ont joué un rôle déterminant tout au long de ce projet.

Je remercie également Monsieur Khaled Ziane pour son encadrement académique et ses conseils tout au long de ce travail.

Ma reconnaissance s'étend aux membres du jury, qui ont bien voulu évaluer ce mémoire et dont les remarques pertinentes ont permis d'en améliorer la qualité.

Je remercie également l'ensemble du corps professoral ainsi que le personnel administratif de l'Université du Québec à Chicoutimi pour leur engagement et leur soutien tout au long de mon parcours.

Un grand merci à mes camarades et amis du Laboratoire d'Automatique et de Robotique interactive (LAR.i) pour leur présence, leur entraide et leur soutien constant.

Ce mémoire a été réalisé au sein du Laboratoire d'Automatique et de Robotique interactive (LAR.i), en partenariat avec le Centre de Recherche et d'Innovation en intelligence énergétique (CR2ie), a utilisé les données provenant de l'Alliance Nationale pour le Big Data dans le domaine des véhicules à énergies nouvelles (NDANEV) et a bénéficié du soutien financier des Fonds de Recherche du Québec – Nature et Technologies (FRQNT), via la bourse de formation en partenariat avec la Régie de l'énergie du Québec (N° 353781) ainsi que la bourse de stage en centre collégial de transfert de technologie (N° 357146).

PLAN DU MÉMOIRE

Ce mémoire est rédigé sous forme d'article scientifique. L'objectif général du travail est d'approfondir la compréhension des interactions entre le style de conduite, les données du véhicule et les conditions de conduite, afin de contribuer à une meilleure estimation de la durée de conduite restante (*Remaining Driving Time*, RDT) et à une gestion plus efficace des systèmes énergétiques des véhicules électriques.

L'article présenté constitue le cœur du mémoire et apporte une contribution scientifique originale en explorant conjointement ces trois composantes. Plusieurs travaux antérieurs ont étudié certains de ces aspects de manière isolée, notamment l'influence de la température ou celle de l'état de charge (SoC), mais peu de recherches ont intégré simultanément les effets combinés du style de conduite, des données dynamiques du véhicule et des conditions de conduite. Ce travail propose ainsi une approche intégrée qui met en évidence les liens entre ces paramètres et leur impact sur la RDT.

La première phase de l'étude est consacrée à la compréhension des interactions entre le style de conduite, les données du véhicule et les conditions de conduite, dans le but d'identifier les paramètres les plus influents sur la performance et la stabilité du système. Dans une seconde phase, ces connaissances sont mobilisées pour le développement d'un modèle prédictif de la RDT, comportant deux volets :

- la prédiction de la durée de conduite entre deux états de charge;
- la prédiction de la durée de conduite jusqu'à la décharge complète;

Ces deux étapes permettent d'approfondir la compréhension des phénomènes influençant la RDT et d'apporter de nouvelles perspectives pour l'optimisation des stratégies de gestion énergétique.

CONTRIBUTIONS DES AUTEURS

L'article a été rédigé principalement par Chourik Fousseni, auteure principale du mémoire. Elle a assuré la conception méthodologique de la recherche, la revue de la littérature, la réalisation des analyses, le développement des modèles prédictifs de la RDT, l'interprétation des résultats et la rédaction intégrale du manuscrit.

Le professeur Martin Otis a contribué à l'orientation scientifique du travail en proposant des idées sur la manière de poursuivre la recherche et en suivant de près l'évolution des analyses afin d'en vérifier la validité. Ses réflexions ont permis d'affiner la structure du travail et d'assurer la cohérence du processus analytique.

Monsieur Khaled Ziane, pour sa part, a apporté un regard critique sur la présentation et l'interprétation des résultats. Ses commentaires ont aidé à améliorer la clarté et la mise en valeur des conclusions. Les échanges avec Martin et Khaled ont permis de définir collectivement les meilleures orientations pour la poursuite du travail et l'enrichissement du contenu scientifique.

Ainsi, la contribution majeure revient à l'auteure principale du mémoire, responsable de la conception, de la mise en œuvre et de la rédaction du projet, tandis que les coauteurs ont apporté un appui scientifique précieux par leurs conseils, leurs réflexions et leurs relectures constructives.

STATUT DU MANUSCRIT

Au moment du dépôt du mémoire, l'article est accepté et publié en ligne dans la revue Machine Learning with Applications (MLWA), publié par Elsevier. L'article est accessible à l'adresse suivante : <https://doi.org/10.1016/j.mlwa.2025.100813>

Le code source du modèle d'apprentissage associé est disponible en ligne à l'adresse suivante : <https://doi.org/10.17605/OSF.IO/D5HF4>

AVANT-PROPOS

L'essor des véhicules électriques s'inscrit aujourd'hui au cœur de la transition énergétique et de la recherche de solutions durables face aux enjeux environnementaux actuels. Dans ce contexte, la maîtrise des performances et de la gestion énergétique des batteries constitue un domaine de recherche essentiel.

Ce mémoire de maîtrise en ingénierie s'inscrit dans cette thématique et a pour objectif d'améliorer la précision de l'estimation du temps de conduite restant des véhicules électriques, en intégrant des paramètres tels que la température et les styles de conduite. Ce travail a été réalisé sous la direction du professeur Martin Otis et la codirection de Monsieur Khaled Ziane, dont l'expertise et les conseils ont grandement contribué à l'avancement de ce projet.

La réalisation de ce projet m'a permis d'approfondir mes connaissances dans les domaines de l'apprentissage automatique, tout en développant une compréhension plus globale des défis techniques associés à l'électrification des transports. Je tiens à exprimer ma gratitude à mes directeurs pour leur accompagnement scientifique, leur disponibilité et leur soutien tout au long de ce travail.

Ce mémoire est structuré en cinq chapitres :

Le chapitre 1 présente l'introduction générale et le contexte de l'étude ;

Le chapitre 2 est consacré à la revue de la littérature, qui expose les principaux travaux liés aux batteries et aux véhicules électriques ;

Le chapitre 3 décrit la méthodologie et les modèles d'apprentissage automatique utilisés ;

Le chapitre 4 présente et discute les résultats obtenus ;

Enfin, le chapitre 5 propose une conclusion générale et des pistes de recherche futures.

CHAPITRE 1

1. INTRODUCTION

Les véhicules électriques (VE) connaissent une croissance remarquable au cours des dernières années. Chatroux, 2021 [1] le souligne clairement dans son étude. En effet, de nombreuses études ont été effectuées pour démontrer la place qu'occupent désormais les VE dans nos sociétés. Notamment, le rôle des VE dans les systèmes de transport et d'énergie a été démontré dans des recherches récentes de Muratori et al., 2021 [2]. Une étude du marché des VE est présentée par Van Mierlo et al., 2021 [3], afin d'estimer leurs utilisations futures et potentielles, et elle démontre le pourcentage mondial estimé d'utilisation des VE en 2025. Cependant, les propriétaires de VE rencontrent plusieurs défis. Les habitudes de conduites diffèrent et les batteries se dégradent avec le temps, entraînant une réduction de leur durée de vie. L'état de santé (*SoH : State of Health*) des batteries des véhicules électriques (BVE) varie selon de nombreux facteurs, dont les cycles de charge et de décharge, les conditions météorologiques, les géométries de la route et les habitudes de conduites (accélération et freinage). Tous ces facteurs sont présentés dans l'étude de El Jebbari, 2021 [4], et l'étude de Rahman & Alharbi, 2024 [5]. De telles conditions pour les batteries des VE entraînent une dégradation inévitable, c'est ce que présente Yapıcı et al., 2025 [6] ainsi que Montes et al., 2024 [7] dans une étude récente. En ce sens, l'utilisation des algorithmes intelligents exploitant les données de conduite réelles offre une solution prometteuse pour prédire la durée de conduite entre deux états de charge et jusqu'à décharge complète de la batterie pour chaque véhicule, selon son utilisation. Autrement dit, la dégradation inévitable des batteries au fil du temps impacte leur autonomie et leur efficacité. Il est donc important de comprendre les différentes habitudes de conduite et leurs impacts sur la batterie afin de mettre en place des stratégies qui permettront à celles-ci d'avoir une efficacité énergétique durable.

La prédiction du temps de décharge d'une batterie d'un véhicule électrique, est un domaine où les algorithmes intelligents peuvent jouer un rôle important, puisqu'ils offrent des solutions pour optimiser la durée de vie des batteries.

Cette étude s'articulera autour de plusieurs phases clés. La première phase consistera à identifier une base de données contenant des données réelles relatives au fonctionnement des véhicules

électriques. La deuxième phase portera sur le nettoyage, la préparation et la sélection des variables pertinentes en vue d'une modélisation efficace. Ensuite, des techniques d'apprentissage automatique seront appliquées pour entraîner des modèles capables de caractériser et de classifier différents profils de conduite. Enfin, ces modèles seront utilisés pour réaliser des prédictions, notamment en ce qui concerne le comportement énergétique du véhicule ou la durée de maintien de la charge, en fonction des profils identifiés. Dans une troisième phase, en se basant sur les analyses précédentes, une analyse sera entreprise pour prédire le temps de maintien de l'état de charge (SoC : *State of Charge*) entre deux états de charge et jusqu'à décharge complète de la batterie.

Des algorithmes, capable d'analyser de grandes quantités de données seront utilisés pour cette prédiction. Cette approche permettra non seulement de prédire le temps de maintien du SoC, mais également de mettre en évidence un lien entre ces différents éléments et de mieux comprendre leur impact sur le SoC.

Dans une quatrième phase, une nouvelle piste sera explorée visant à utiliser la validation croisée pour la prédiction du temps de conduite entre deux états de charge.

1.1. Contexte

Contrairement aux propriétaires des véhicules à essence, les propriétaires des véhicules électriques sont confrontés à un nouvel enjeu: la gestion de la durée de vie et de la performance de leur batterie. Effectivement, avec les conditions météorologiques, les styles de conduite, le système de freinage, l'accélération, les cycles de charge et de décharge, les batteries se dégradent progressivement au cours de leur utilisation.

Pour relever ce défi, l'utilisation d'un algorithme intelligent, basé sur des données d'utilisation en contexte réel, se présente comme une solution prometteuse. À ce jour et à notre connaissance, aucune étude n'a abordé l'utilisation des données réelles pour estimer le temps de maintien de la charge entre un état de charge de départ et un état de charge d'arrivée. En analysant les caractéristiques d'habitude d'utilisation ainsi que l'état de charge des batteries, la prédiction du

temps de maintien du SoC pourrait être utilisée à différentes fins, notamment celle de l'adaptation de la conduite automobile.

Dans ce même cadre, les algorithmes intelligents seront employés pour établir des relations quantitatives entre les paramètres de conduite, tels que la vitesse et la consommation énergétique, et l'autonomie restante; avec un regroupement des profils de conduite similaires afin d'identifier les tendances d'utilisation. Des techniques plus avancées, comme les algorithmes temporels, et les modèles de gradient boosting permettront d'améliorer la précision des prédictions et de trouver une corrélation entre les différentes variables.

1.2. Problématique

La gestion de la performance et de la durée de vie des batteries des véhicules électriques constitue un défi de taille dans le domaine de la mobilité électrique. En effet, cette gestion nécessite une intégration des variables complexes telles que le SoC, les conditions météorologiques, et les styles de conduite. Cette problématique soulève la question de l'établissement d'un lien entre toutes les variables liées aux VE et le temps de maintien du SoC.

Comprendre la façon dont ces variables interagissent et influencent la performance globale des batteries est essentiel pour développer des stratégies de gestion de la performance. Toutefois, la problématique ne se limite pas uniquement à la gestion des batteries des véhicules électriques en service. Elle s'étend également à l'obtention d'un lien entre toutes les variables, notamment la tension, et le courant de la batterie et la vitesse de conduite. Cette problématique nécessite de vérifier comment le temps de maintien du SoC peut être liée à toutes les autres variables.

1.3. Objectifs

Afin de développer un système de gestion pour les batteries, autant pour les VE que pour d'autres applications, il est possible d'intégrer des algorithmes intelligents basés sur les données d'utilisation en contexte réel afin de prédire le temps de maintien du SoC selon le comportement d'utilisation passée. L'objectif principal de cette étude consiste à établir un lien entre le SoC, et le temps restant de conduite en fonction des habitudes d'utilisation en entraînant cinq (5)

algorithmes, *Extreme Gradient Boosting Regression Tree (XGBoost)*, *Gradient Boosting Regression Tree (LightGBM)*, *Long Short- Term Memory (LSTM)*, *Random Forest (RF)* et *Prophet*, permettant ainsi de prédire la durée de maintien de la charge pour une batterie chargée à un certain pourcentage. Cette approche aspire à fournir une fiabilité et une gestion durable de l'énergie dans divers contextes, y compris les applications de conduite automobile.

La figure 1 illustre l'ensemble des variables influençant la performance de la batterie d'un VE. Cette figure présente les relations entre différents facteurs tels que le SoC, les habitudes de conduite, les conditions de conduite permettant de faire une estimation du temps restant de conduite d'un VE.

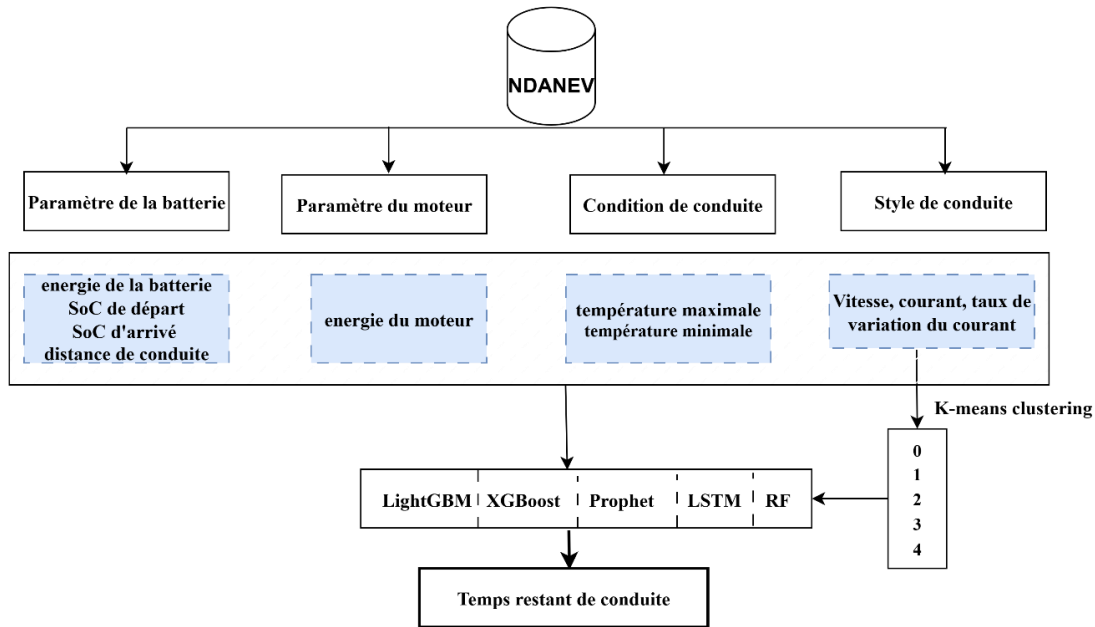


Figure 1: Interaction entre les variables influençant la performance de la batterie d'un VE

1.4. Contribution

Après avoir passé en revue la littérature existante, une analyse comparative des travaux pertinents est effectuée dans le tableau 1. Dans de nombreuses recherches, le lien entre le SoC et les conditions météorologiques a été établi. Cependant, jusqu'à présent et selon notre analyse bibliométrique, aucun lien entre le SoC, le style de conduite et le temps de maintien du SoC en fonction des conditions de conduite n'a été trouvé. Pourtant, il est connu que le temps de maintien

du SoC est influencé par divers facteurs, notamment l'état de la batterie, l'état de la charge et les conditions de conduite. Dans cette étude, l'établissement d'un lien entre le SoC, l'énergie et le temps de maintien du SoC en fonction des conditions de conduite constitue la contribution principale à cette recherche.

L'étude s'étend également à la gestion des relations quantitatives entre les variables de conduite, tels que la vitesse et la consommation énergétique. Cette relation quantitative représente une autre facette de la gestion des batteries.

Table 1: Comparaison des travaux de recherche antérieurs selon leurs contributions

Article	Modèles	Variables	Objectif
Notre approche	Prophet, RF, LSTM, XGBoost, LightGBM	Temperature, battery_energy, motor_energy, drive_distance, start_soc, end_soc, used_soc, group	Estimation du temps de conduite entre deux états de charge, en tenant compte du style de conduite et des conditions de conduite
L. Zhao et al., 2020 [8]	LightGBM, XGBoost, Modèle mixte	Speed, used_soc, motor_energy, battery_energy, drive_distance, time, temp max, temp min, group	Prédiction de la distance de conduite en tenant compte du style de conduite et de la température
N. Wang et al., 2024 [9]	TCN, GRU	Speed, battery voltage, battery energy, mileage, ambient_temp	Prédiction de l'autonomie restante de la batterie à partir de l'énergie de décharge restante et du taux de consommation d'énergie
Acharyaviriya et al., 2023 [10]	XGBoost, RF, MLP, SVR	SoC, battery voltage, battery_current, motor_current, acceleration, road_gradient	Estimation de la consommation d'énergie des véhicules électriques à batterie à l'aide d'approches d'apprentissage automatique
Eissa & Chen, 2023 [11]	CNN-LSTM	speed, SoC, energy_consumed,	Une approche hybride efficace d'apprentissage profond pour une

Article	Modèles	Variables	Objectif
		ambient_temp	prédiction précise de l'autonomie restante des véhicules électriques

1.5. Organisation des chapitres

Ce mémoire est structuré en cinq chapitres, chacun abordant une étape clé du projet de recherche :

Le chapitre 1 introduit le projet en le situant dans son contexte général. Il y est présenté la problématique à laquelle cette étude tente de répondre ainsi que les objectifs spécifiques poursuivis.

Le chapitre 2 est consacré à une revue de littérature portant sur les travaux antérieurs relatifs à l'estimation du temps de conduite d'un véhicule électrique, ce qui permet de positionner le projet par rapport aux approches existantes.

Les chapitres 3 et 4 forment le cœur du mémoire. Ils présentent la méthodologie mise en œuvre ainsi que les résultats obtenus. Il s'agit, à ce niveau, de l'article scientifique structurant ce travail, tel qu'il a été soumis.

Le chapitre 5 propose une synthèse des résultats sous forme de conclusion, et expose les perspectives envisagées pour des travaux futurs dans la continuité de cette recherche.

CHAPITRE 2

2. REVUE DE LA LITTÉRATURE

Dans le cadre du projet de recherche portant sur l'estimation du temps de maintien du SoC via l'utilisation d'algorithmes intelligents adaptés aux profils d'utilisation, une revue de la littérature est entreprise afin de comprendre les avancées récentes dans ce domaine. Cette revue de la littérature permet d'identifier les recherches antérieures pertinentes et de comprendre les différentes approches utilisées pour situer la portée de ce travail.

Huang et al., 2020 [12] ont réalisé une estimation de l'état de charge (SoC) des batteries en fonction des habitudes d'utilisation des véhicules électriques (VE), en se basant sur les données collectées par la *National Big Data Alliance of New Energy Vehicles* (NDANEV). Ces données, provenant de cinq véhicules électriques suivis sur une période d'un an, ont permis de générer environ un million d'enregistrements après nettoyage. Pour surmonter les défis liés à la complexité et à la discontinuité des données de kilométrage et de SoC, Huang et al., 2020 [12] ont développé deux algorithmes spécifiques. Le premier, l'algorithme de régression basé sur l'interpolation du SoC, et le second, l'algorithme de classification basé sur l'accumulation de conduite, prennent en compte les habitudes de conduite, modifient les données de kilométrage pour les rendre continues, corrigent les positions géographiques (longitude et latitude), et segmentent les conduites continues en fonction du temps. Ces améliorations permettent d'obtenir une estimation du SoC plus précise, essentielle pour la gestion des performances des batteries lithium-ion dans les véhicules électriques.

Dans une étude récente, Min et al., 2023 [13] ont proposé une méthode basée sur la méthode du poids entropique pour estimer l'état de santé (SoH) de la batterie, en utilisant des données de conduite réelles. Leur approche prend en compte des paramètres essentiels tels que la capacité kilométrique qui représente la distance qu'une voiture parcourt avec une charge complète de batterie, le temps de charge et la température. En exploitant ces données, ils ont développé une approche permettant d'estimer le SoH de la batterie, offrant ainsi aux propriétaires des véhicules électriques un outil crucial pour évaluer la performance et la durabilité de la batterie.

Par ailleurs, Zhao et al., 2020 [8] ont proposé une méthode innovante pour prédire l'autonomie restante des véhicules électriques, un enjeu majeur en raison de l'autonomie limitée qui freine leurs adoptions. Ce modèle d'apprentissage mixte s'appuie sur des données de conduite réelles et combine deux algorithmes avancés, à savoir le *Extreme Gradient Boosting Regression Tree (XGBoost)* et le *Light Gradient Boosting Regression Tree (LightGBM)*. La formation du modèle se concentre sur la relation entre la distance de conduite et des caractéristiques clés telles que l'énergie de sortie cumulée du moteur et de la batterie, les divers styles de conduite, ainsi que la température de la batterie. Une stratégie basée sur l'ancrage de base a également été introduite pour corriger la distribution déséquilibrée des données.

Les batteries lithium-ion sont largement adoptées pour alimenter les véhicules électriques, selon Ghalkhani & Habibi, 2023 [14]. L'estimation de l'état des batteries est cruciale pour optimiser leurs efficacités énergétiques, prolonger leurs durées de vie et assurer une intégration sûre dans les VE, l'étude de J. Liu et al., 2024 [15] en fait une analyse approfondie. Cependant le caractère non linéaire du comportement des batteries, ainsi que la variation temporelle de certains compliquent l'estimation des paramètres clés tels que la résistance interne, le SoC et le SoH. Cette estimation dépend de divers facteurs tels que l'âge et la température des batteries. Une classification des méthodologies d'estimation des paramètres internes des batteries est essentielle pour sélectionner les approches les plus adaptées au développement de système de gestion des batteries, comme le souligne l'étude de Laadjal & Cardoso, 2021 [16].

En outre, Hong et al., 2021 [17] avaient pour objectif d'analyser comment la consommation énergétique des VE urbains varie en fonction des différentes régions d'exploitation, des échelles de temps et des types de groupes de VE. L'analyse a été menée sur deux catégories : les VE publics et les VE privés, et montre comment les variations saisonnières influencent le comportement des utilisateurs vis-à-vis des batteries. Une méthode similaire à celle de l'étude précédente a été utilisée, tout en intégrant les données de température dans leur analyse. Des simulations ont été réalisées, démontrant que la prise en compte de ces facteurs dans le développement d'algorithmes de gestion de la batterie peut permettre de mieux comprendre et de limiter la dégradation sous des conditions réelles d'utilisation.

Une contribution récente de N. Wang et al., 2024 [9] propose une approche hybride combinant des modèles basés sur les données et des modèles physiques pour prédire le temps de conduite restante (RDT) des véhicules électriques. La RDT est décomposée en deux composantes clés : l'énergie restante à décharger (RDE) et le taux de consommation d'énergie (ECR), lesquels sont ensuite liés à l'état de santé (SoH) et à un coefficient de consommation ECR. Ces paramètres sont prédits à l'aide d'un modèle LightGBM et d'un modèle TCN-GRU amélioré. L'approche a été validée à partir de données réelles collectées sur cinq taxis électriques à Shanghai pendant une période de 30 mois. Bien que l'étude prenne en compte diverses conditions de conduite telles que la vitesse et la température elle ne considère ni les styles de conduite individuels ni les conditions de conduite dynamiques pour prédire le temps réel de conduite. Le présent projet répond précisément à cette lacune.

Une autre étude pertinente de Achariyaviriya et al., 2023 [10] porte sur l'estimation de la consommation d'énergie dans des conditions réelles de conduite, en utilisant des données de capteurs recueillies via le système OBD-II. Les auteurs ont analysé environ 35 000 trajets courts provenant de plusieurs BEV circulant en Thaïlande. L'ensemble de données comprenait des variables telles que la vitesse du véhicule, l'accélération, le courant de la batterie, l'inclinaison de la route, l'état de charge et la position GPS, enregistrées à une fréquence de 1 Hz. Quatre algorithmes d'apprentissage automatique, XGBoost, Forêt aléatoire (Random Forest), Perceptron multicouche (MLP) et Régression par vecteurs de support (SVR) ont été utilisés pour modéliser la consommation d'énergie, avec une évaluation des performances par validation croisée à dix plis. Un pipeline de prétraitement rigoureux a été appliqué, incluant la normalisation des caractéristiques et la transformation non linéaire de Yeo-Johnson, afin d'améliorer la stabilité de l'apprentissage. Les résultats, analysés avec la méthode SHAP (*SHapley Additive exPlanations*), ont mis en évidence le courant de la batterie et la vitesse du véhicule comme les variables les plus influentes sur la consommation d'énergie. Cette étude apporte des informations sur l'usage réel de l'énergie et identifie les principaux facteurs d'influence, contribuant à une meilleure efficacité énergétique.

Contrairement à l'approche développée dans le présent travail, l'étude Eissa & Chen, 2023 [11] propose un modèle hybride combinant un réseau de neurones convolutifs (CNN) et un réseau de type *Long Short-Term Memory* (LSTM) pour prédire l'autonomie restante des véhicules électriques à partir de données historiques de conduite. Le modèle est construit à partir de données réelles recueillies sur une Nissan Leaf 2019 conduite par plusieurs utilisateurs. Les caractéristiques utilisées incluent l'état de charge (SoC), le courant et la tension de la batterie, la température de la batterie, la vitesse du moteur, la vitesse du véhicule et la température ambiante. Après normalisation, les données sont structurées en séquences temporelles à l'aide de fenêtres glissantes. Le CNN extrait automatiquement les caractéristiques pertinentes, qui sont ensuite transmises au LSTM pour effectuer les prédictions. L'ensemble de données a été divisé en deux sous-ensembles : 90 % pour l'entraînement et 10 % pour les tests. La capacité de généralisation du modèle a été évaluée à l'aide de données issues de nouveaux utilisateurs n'ayant pas participé à la phase d'entraînement.

Il est important de souligner que la plupart des études antérieures dans ce domaine se sont principalement concentrées sur l'analyse des données des véhicules en prenant en compte des facteurs tels que le modèle, le type de batterie ou la puissance de la batterie, ainsi que les conditions météorologiques. Cependant, ces données ne fournissent qu'une partie de la complexité de la gestion de l'autonomie de la batterie des véhicules électriques. En effet, le SoC, l'énergie et le temps de maintien du SoC en fonction des habitudes d'utilisation exercent une influence significative sur l'autonomie de la batterie. Par conséquent, une approche mettant en lien le SoC et le temps de maintien du SoC en fonction des habitudes d'utilisation est nécessaire pour optimiser l'autonomie des véhicules électriques.

À la suite de cette revue de littérature, une taxonomie a été établie pour classer et organiser les éléments identifiés. La figure 2 présente cette synthèse, regroupant les principales catégories et sous catégories afin de faciliter la compréhension des différentes approches recensées.

2.1. Définitions et contextualisation des concepts

Avant de présenter les aspects techniques de l'étude, une contextualisation des principaux concepts a été réalisée. Cette démarche permet de clarifier les notions clés mobilisées, depuis la

collecte des données jusqu'à l'évaluation des performances. Chaque terme a été défini en fonction de son rôle et de sa pertinence dans le domaine d'application, conformément aux définitions proposées par L. Zhao et al., 2020 [8].

2.1.1. SoC

Le *State of Charge* (SoC) correspond à la quantité d'énergie restante dans la batterie d'un véhicule électrique, exprimée en pourcentage de la capacité nominale totale. Il constitue un indicateur essentiel pour estimer l'autonomie restante et planifier les recharges. Dans cette étude, le SoC est utilisé comme variable centrale pour évaluer le temps pendant lequel la batterie peut conserver un certain niveau de charge, en fonction de différents facteurs.

2.1.2. SoH

Le *State of Health* (SoH) évalue l'état général de la batterie par rapport à ses performances initiales. Il est généralement exprimé en pourcentage de la capacité nominale d'origine. Un SoH réduit indique une dégradation de la batterie, ce qui impacte directement l'autonomie et la sécurité du système. Le SoH permet ainsi d'intégrer une dimension temporelle à l'analyse, en tenant compte du vieillissement des batteries dans les prédictions.

2.1.3. Style de conduite

Les habitudes de conduite influencent considérablement la consommation d'énergie. Des variables telles que la vitesse moyenne, la variation du courant moteur, le type de trajet (urbain, autoroute) ont été prises en compte pour caractériser le style de conduite. Ces éléments permettent d'identifier des profils de conducteurs et de les relier aux performances énergétiques observées.

2.1.4. Conditions de conduite

Les conditions environnementales jouent également un rôle déterminant. La température extérieure, la géolocalisation de départ, et parfois l'altitude influencent l'efficacité énergétique du véhicule. Par exemple, des températures basses réduisent la performance des batteries

lithium-ion. Dans cette étude, ces conditions sont intégrées pour affiner les estimations du SoC et mieux comprendre les variations observées dans les données.

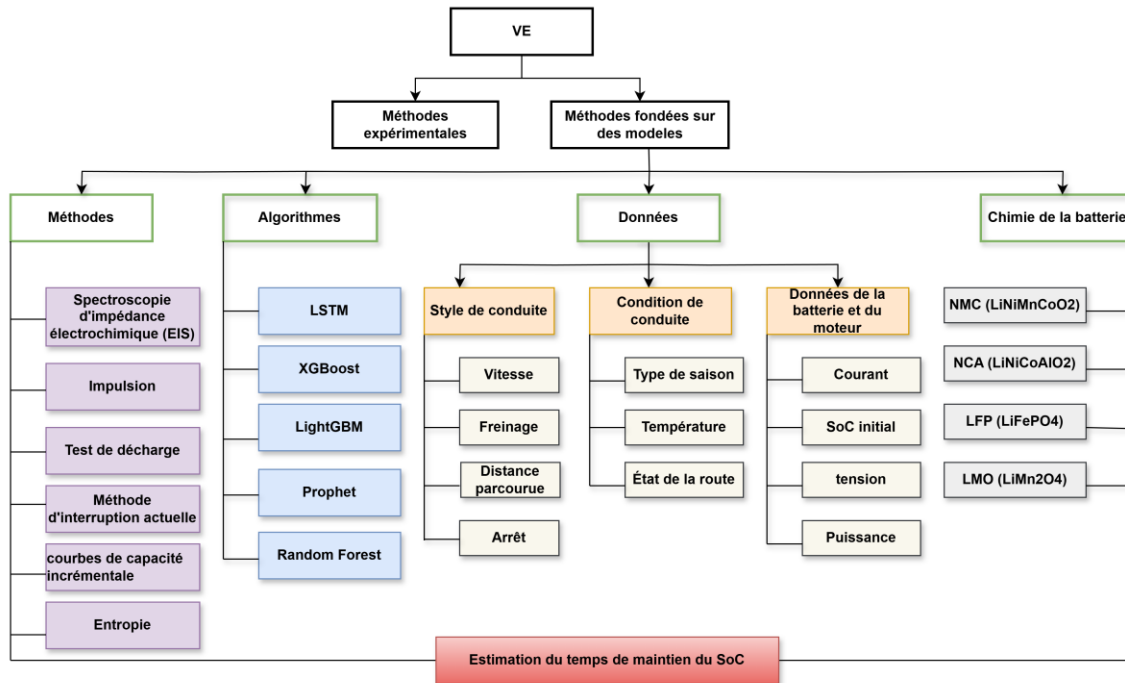


Figure 2: Taxonomie des approches et algorithmes

2.2. Analyse bibliométrique

L'analyse bibliométrique est une étape essentielle de toute recherche. Elle permet de comprendre les tendances, d'identifier les auteurs influents et de connaître les travaux antérieurs réalisés dans un domaine spécifique. Cette analyse réalisée par VOSviewer qui est un outil de recherche scientifique avancé, pour explorer les tendances et les contributions dans tout domaine, permettra de situer cette étude et de fournir des informations orientant la recherche.

Des mots clés spécifiques ont été utilisés dans l'outil Scopus afin d'identifier les travaux pertinents dans le domaine de l'optimisation de l'autonomie des véhicules électriques. Parmi ces mots clés, "*State of Health*" (SoH) et "*State of Charge*" (SoC) ont été inclus, identifiant l'état de santé et l'état de charge de la batterie, ainsi que "*Electric car*" ou "*Electric vehicle*" pour orienter l'étude vers les véhicules électriques, "*Battery*" pour parler du composant principal des

Groupe 1 (rouge) : centré sur la gestion des batteries et des cycles de charge, regroupant des termes comme *charging* (batteries), *electric vehicle* (EV), et *battery management system* (BMS).

Groupe 2 (vert) : orienté vers la santé, incluant *state of health* (SoH), *state of charge* (SoC), et *health*.

Groupe 3 (bleu) : associé aux approches d'apprentissage automatique et d'optimisation, avec des mots-clés tels que *machine learning*, *learning systems*, *errors* et *mean square error*.

Groupe 4 (jaune) : relatif à la modélisation, intégrant *estimation methods*, et *accurate estimation*.

Groupe 5 (violet) : focalisé sur les aspects chimiques, notamment *lithium-ion battery*, *lithium ions* et *ion batteries*.

Groupe 6 (bleu ciel) : lié aux systèmes de stockage et de décharge, incluant *depth of discharges*.

2.3. Analyse des données utilisées dans les études précédentes

Les données des véhicules électriques, des conditions d'utilisation, et des conditions de conduite permettront à l'aide des algorithmes d'établir un lien entre le SoC, les habitudes d'utilisation et le temps de maintien du SoC. L'examen des différentes bases de données utilisées dans les recherches antérieures est essentiel pour déterminer les types de données appropriées à utiliser.

Le tableau 1 présente un récapitulatif des principales bases de données exploitées dans les travaux précédents. Ces bases de données contiennent des informations pertinentes sur les véhicules électriques, les données de conduite, ainsi que les conditions de conduite.

Cette analyse comparative assure une approche méthodologique, favorisant l'intégration des données pertinentes pour cette étude, facilitant ainsi la sélection des bases de données pertinentes pour le projet en cours.

En s'appuyant sur les informations recueillies, il est possible d'élaborer des stratégies de gestion des batteries plus efficaces et de développer des modèles prédictifs pour améliorer la

performance et la durabilité de la batterie des véhicules électriques. De plus, cette recherche contribue à l'avancement des connaissances sur les dynamiques de charge et de santé des batteries, offrant des perspectives nouvelles pour les futures études dans ce domaine.

Le choix des données s'est porté sur celles présentant la plus longue période de collecte et provenant d'études les plus proches du contexte de la présente recherche. L'étude L. Zhao et al., 2020 [8] a notamment servi de référence pour orienter et justifier cette sélection.

RÉSUMÉ ÉTENDU

L'estimation précise du temps de conduite restant (RDT) de la batterie d'un véhicule électrique (VE) est essentielle pour optimiser la gestion de l'énergie et améliorer l'expérience utilisateur. Cependant, les méthodes d'estimation traditionnelles ne prennent pas suffisamment en compte l'influence de la température, des caractéristiques de conduite et du temps de conduite du véhicule, ce qui conduit à des prédictions moins précises et à une gestion de l'autonomie sous-optimale. Afin de remédier à ces limitations, cette étude présente une méthode d'estimation du temps restant de conduite en intégrant la température et les caractéristiques de conduite, ce qui permet d'affiner les prédictions et d'améliorer la fiabilité du modèle. En outre, des données issues de la National Big Data Alliance for New Energy Vehicles (NDANEV) ont été utilisées pour développer un modèle prédictif basé sur des modèles d'apprentissage automatique (ML). Les différents modèles de ML comparés dans cette étude sont le LSTM, le RF, Prophet, LightGBM et XGBoost. Les performances des modèles ont été évaluées à l'aide de l'erreur absolue moyenne (MAE), de la racine de l'erreur quadratique moyenne (RMSE), du coefficient de détermination (R^2) et du temps d'exécution de la prédiction afin d'évaluer les prédictions. Les résultats montrent que les valeurs de R^2 pour Prophet, Random Forest, LSTM, XGBoost et LightGBM sont respectivement de 0,91, 0,94, 0,95, 0,94 et 0,94. Cela suggère que XGBoost surpasse les autres modèles en fournissant l'estimation la plus précise du temps de conduite restant. De plus, les résultats confirment que la prise en compte des caractéristiques de conduite et de la température ambiante améliore la fiabilité des estimations. Ces avancées pourront contribuer à une gestion de l'énergie plus efficace et à des stratégies de recharge optimisées.

Mots-clés : Véhicule électrique (VE) ; état de charge (SoC) ; conditions et caractéristiques de conduite ; batterie lithium-ion (LIB) ; apprentissage automatique (ML) ; Random Forest (RF) ; Long Short-Term Memory (LSTM), Prophet, XGBoost, LightGBM.

Estimation of the Remaining Charge Retention Time of an Electric Vehicle Battery

CHAPITRE 3

MÉTHODOLOGIE

Chourik Fousseni ^{a*}, Martin Otis ^a, Khaled Ziane ^b

^a Autonomous and interactive Robotic Lab, Université du Québec à Chicoutimi, Québec, G7H 2B1, Canada

Emails: cfousseni@uqac.ca, m3otis@uqac.ca

^b Centre de Recherche et d'Innovation en Intelligence Énergétique CR2ie, 175 Rue de la Vérendrye, Sept-Îles, QC G4R 5B7, Canada

Email: khaled.ziane@cegepsi.ca

* Author to whom correspondence should be addressed

ABSTRACT

Accurately estimating the remaining driving time (RDT) of an electric vehicle (EV) battery is essential for optimizing energy management and enhancing user experience. However, traditional estimation methods do not adequately account for the influence of temperature, driving characteristics and vehicle driving time, leading to less accurate predictions and suboptimal range management. To address these limitations, this study presents a method for estimating the remaining charge retention time by integrating temperature and driving characteristics, which refines predictions and improves model reliability. Furthermore, data from the National Big Data Alliance for New Energy Vehicles (NDANEV) were employed to develop a predictive model based on machine learning (ML) models. The different ML models compared in this study are Linear Regression, LSTM, RF, Prophet, LightGBM, and XGBoost. The model performance was evaluated using mean absolute error (MAE), root mean square error (RMSE), the coefficient of determination (R^2) and the prediction runtime to assess the prediction accuracy. The results show that the R^2 values for Prophet, Random Forest, LSTM, XGBoost, and LightGBM are 0.91, 0.94, 0.95, 0.94, and 0.94 respectively. This suggests that XGBoost outperforms the other models, providing the most accurate estimate of the remaining driving

time. In addition, the result confirms that considering driving characteristics and ambient temperature improves the reliability and robustness of estimations. These advancements contribute to more efficient energy management and optimized charging strategies.

Keywords: Electric vehicle (EV); state of charge (SoC); driving conditions and characteristics; Lithium-ion Battery (LIB); machine learning (ML); Random Forest (RF); Long Short-Term Memory (LSTM).

INTRODUCTION

The growing adoption of electric vehicles (EVs) stems from their potential to enhance energy efficiency and mitigate environmental impacts as reported by Kapustin & Grushevenko, 2020 and Pelegov & Chanaron, 2023 [18, 19]. As their role in transportation and energy systems expands, numerous studies have emphasized their increasing importance in modern societies, as highlighted by Muratori et al., 2021 [2]. Driving conditions play a crucial role in estimating an EV's remaining driving time (RDT). External factors such as ambient temperature, weather conditions, and topography significantly influence battery energy consumption as demonstrated by Li et al., 2025 [20]. In addition, driving behavior also has a strong impact on energy consumption, directly affecting the RDT.

Despite advances in battery management and vehicle monitoring systems, accurately estimating the remaining discharge time of EV batteries remains a major challenge. Existing models are often limited by multiple factors, such as driving conditions, driver behavior, and battery degradation over time, as highlighted by Masakure et al., 2023 [21]. These limitations result in unreliable estimates of remaining range and create uncertainty for drivers regarding battery depletion. Moreover, most studies focus primarily on predicting the state of charge (SoC) or remaining driving distance, while the prediction of RDT a distinct concept referring to the duration between two successive charge states has received limited attention. Consequently, there is a need for models that integrate heterogeneous factors, including driver behavior, environmental conditions, and energy dynamics, to improve prediction accuracy.

For example, consider an electric vehicle traveling 10 km under moderate traffic with an economical

driving style and an ambient temperature of 25 °C. The proposed model can estimate the corresponding driving time based on these key factors, illustrating the practical application of the study.

To fill these gaps, the present study proposes a machine-learning-based framework for accurately estimating EV remaining driving time. The methodology incorporates data preprocessing steps, including anomaly removal, missing-value interpolation, feature engineering, and normalization, followed by model training and evaluation. The approach also considers different driving styles by creating five distinct driver groups, allowing the models to adapt predictions to various driver behaviors. Ensemble algorithms such as LightGBM and XGBoost are compared with other machine learning models, including Prophet, RF, and LSTM, to identify the optimal balance between predictive accuracy and computational efficiency.

The main contributions of this study are as follows:

- Definition and formalization of the RDT concept as a temporal measure of EV operation, distinct from range estimation.
- Development of a comprehensive machine-learning framework that integrates vehicle, and environmental parameters for RDT prediction.
- Creation of five driver groups to account for different driving styles and improve prediction accuracy.
- Comparative evaluation of five predictive algorithms, analyzing trade-offs between accuracy and computational cost.
- Analysis of the influence of driver behavior and external factors, such as temperature variations, on model performance.

The remainder of this paper is organized as follows. Section 2 presents a comprehensive review of related work and identifies the research gaps. Section 3 details the proposed methodology and data-processing pipeline. Section 4 discusses the experimental results and model comparisons. Finally, section 5 concludes the paper and outlines potential directions for future research.

RELATED WORKS

Electric vehicles (EVs) are increasingly being adopted worldwide as part of the broader shift toward sustainable transportation and the reduction of greenhouse gas emissions. However, this growing adoption brings new challenges, particularly in energy management. The estimated driving times shown on EV dashboards are often inaccurate, as they fail to incorporate various external and contextual factors such as individual driving behavior and weather conditions. These limitations can hinder drivers' decision-making and trip planning, especially when charging infrastructure is scarce or unevenly distributed. The present project addresses this issue by proposing a more accurate method for predicting real driving time from a given state of charge, considering both driving behavior and external driving conditions. To better position our project within the current scientific landscape, a comparative table of existing articles has been created. While table 2 focuses on the methodological aspects of the reviewed studies, table 3 highlights their main advantages and limitations.

L. Zhao et al., 2020 [8] proposed a hybrid model combining LightGBM and XGBoost is used to predict the motor energy consumption of an electric vehicle, to estimate the remaining driving distance until the battery reaches a minimum SoC threshold. The model is trained and tested on real-world data from 120 driving trips, using driving parameters such as speed and motor current. In contrast, our approach aims to directly predict the remaining time until the minimum SoC threshold is reached, avoiding the intermediate energy estimation step. This prevents error accumulation from sequential predictions and provides a more direct and potentially more reliable estimation under real-world conditions.

A recent contribution proposes a hybrid approach that combines data-driven and physics-based models to predict the remaining driving range (RDR) of electric vehicles, as proposed by Wang et al., 2024 [9]. The RDR is decomposed into two key components: remaining discharge energy (RDE) and energy consumption rate (ECR), which are then linked to the state of health (SoH) and a consumption coefficient $ECR \propto$. These parameters are predicted using a LightGBM model and an enhanced TCN-GRU model. The approach was validated using real-world driving data collected from five electric taxis in Shanghai over a period of 30 months. While the study accounts for various real conditions such as speed, temperature, and data sampling frequency, it does not consider individual driving

styles nor dynamic driving conditions in predicting the actual driving time. The present project addresses this specific gap.

Acharyaviriya et al., 2023 [10] conducted a relevant study focusing on the estimating of energy consumption in battery electric vehicles (BEVs) under real-world driving conditions using sensor data collected through the OBD-II system. The authors analyzed approximately 35,000 short-trip data points from multiple BEVs operating in Thailand. The dataset included variables such as vehicle speed, acceleration, battery current, road gradient, state of charge, and GPS position, all recorded at a frequency of 1 Hz. Four machine learning models XGBoost, Random Forest, Multilayer Perceptron (MLP), and Support Vector Regression (SVR) were employed to model energy consumption, with performance evaluated through tenfold cross-validation. A robust preprocessing pipeline was applied, including feature normalization and the Yeo-Johnson nonlinear transformation, to enhance model learning stability. The results further analyzed using the SHAP (SHapley Additive exPlanations) approach, highlighted battery current and vehicle speed as the most influential variables affecting energy consumption. The study provides valuable insights into the real-world energy usage of BEVs and identifies key influencing factors, contributing to improved energy efficiency and supporting policy decisions toward transportation decarbonization.

Unlike the approach developed in the present work, Eissa & Chen, 2023 [11] propose a hybrid model combining a Convolutional Neural Network (CNN) and a Long Short-Term Memory (LSTM) network to predict the remaining driving range of electric vehicles based on historical driving data. The model is built using real-world data collected from a 2019 Nissan Leaf driven by multiple users. The features used include the SOC, battery current and voltage, battery temperature, motor speed, vehicle speed, and ambient temperature. After normalization, the data are structured into time sequences using sliding windows. The CNN automatically extracts relevant features, which are then passed to the LSTM to make predictions. The dataset was split into two subsets: 90% for training and 10% for testing. The model's generalization capability was assessed using data from new users who were not involved in the training phase.

Table 2: Comparison of our results with existing literature works

Article	Method	Features	Objectives
Our Approach	Prophet, RF, LSTM, XGBoost, LightGBM	Temperature, battery energy, motor energy, driving distance, start_soc, end_soc, used_soc, group	Estimation of driving time between two charge states, considering driving style and driving conditions
[8]	LightGBM, XGBoost, Blended model	Speed, used_soc, energy of the motor, energy of the battery, driving time, temp max, temp min, group	Prediction of driving distance considering driving style and temperature
[9]	TCN, GRU	Speed, battery voltage, battery energy, mileage, temperature	Prediction of remaining battery life using remaining discharge energy and energy consumption rate
[10]	XGBoost, RF, MLP, SVR	SoC, battery voltage, battery current, acceleration, road gradient	Estimating the energy consumption of battery electric vehicles using machine learning approaches
[11]	CNN-LSTM	Speed, SoC, energy consumption, ambient temperature	An Efficient Hybrid Deep Learning Approach for Accurate Remaining EV Range Prediction

Table 3: Comparison of existing techniques: advantages and limitations

Articles	Advantages	Limits
[8]	Consider driving styles and ambient temperature.	Does not perform normalization processes
[9]	Takes temperature into account and normalizes data	Does not take driving style into account
[10]	Consider the condition of the road and normalize the data	Does not consider driving style and ambient temperature
[11]	Use time models and take temperature into account	Does not take driving style into account

3. METHODOLOGY

The flowchart in figure 4 illustrates the methodology employed in this study, outlining the key steps and processes involved in achieving the stated objectives. The methodology is organized into several main phases: data collection (section 3.1), data analysis (section 3.2), k-means clustering (section 3.3), RDT predictions models with justifications (section 3.4), cross-validation (section 3.5),

hyperparameter selection (section 3.6), performance metrics selection (section 3.7) and statistical tests (section 3.8). At each stage, appropriate tools and techniques were employed to ensure data quality and analytical rigor. Therefore, the next subsection introduces the initial phases, beginning with the collection of data related to electric vehicle usage, followed by an in-depth analysis and selection of the variables to be used for model training. This process ultimately leads to the prediction stage, where the developed models generate charge retention time estimates based on the selected features.

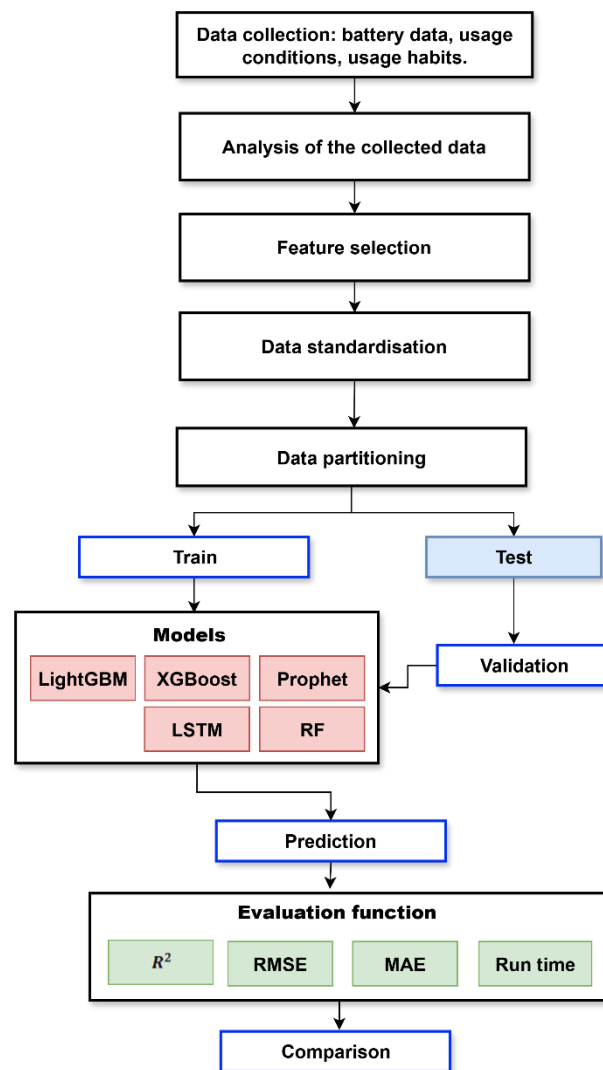


Figure 4: Flowchart for methodology used

3.1. Data collection

The dataset used in this study comes from the NDANEV alliance, which contains a large volume of real-world EV driving data. The NDANEV alliance is a national, non-profit social organization formed by various stakeholders in the new energy vehicle sector in China, aimed at collecting and analyzing driving data to improve the technology and performance of electric vehicles, as described by Xu & al., 2023 [22].

To provide a clear overview, all information related to the vehicle type, model, and manufacturer is presented in table 4.

Table 4: Summary of Vehicle Characteristics and Collection Period

Data information	Correspondences	Details
Data collection platform	NDANEV	https://www.ndanev.com/
Data collection period	2018-01-01 to 2019-12-31	24 months
Vehicle type	Battery Electric Vehicle (BEV)	Fully electric vehicle
Manufacturer	Build Your Dreams (BYD)	Chinese manufacturer
Vehicle model	EV200	Electric Vehicle 200
Motor type	TZ200XSQ02	Electric motor
Battery family	Lithium-ion	Lion
Chemical type	Lithium Iron Phosphate	LiFePO ₄
Data collection zone	Shanghai, Beijing	City in China

This database includes crucial information that helps capture the driving profile of EV, which is essential for improving battery performance, energy consumption, and vehicle range. As summarized in table 5, the dataset offers valuable insights into various parameters such as temperature, used_soc, driving distance and other environmental factors that impact EV performance. The symbols and notations used for these variables throughout this work are summarized in table 6.

The estimation of driving time between an initial and a final state of charge relies on energy quantities calculated from raw data. The mechanical energy produced by the motor is determined from the maximum power and the trip duration, according to the formula (1).

$$Motor_energy = \frac{\max_power * time}{3600}, \quad (1)$$

where max_power is expressed in kilowatts (KW), and time = ed_time-st_time represents the trip duration in seconds. Dividing by 3600 converts the motor energy into kilowatt-hours (KWh).

On the battery side, the energy drawn is calculated from the maximum voltage, current, and duration using the formula (2).

$$Battery_energy = \frac{\max_voltage * \max_current * time}{3600 * 1000}, \quad (2)$$

where max_voltage is in volts, max_current in amperes, and time is seconds. The numerator gives energy in watt-seconds and dividing by 3600*1000 converts watt-seconds (or joules) to kilowatt-hours.

These two energy quantities, expressed in Kilowatt-hours, serve as the main input variables used to train and feed the model predicting driving time between the initial and final states of charge.

These two quantities serve as input variables for the model.

The initial phase focuses on training the model. Raw data including battery parameters, and driving logs, enable the model to learn the relationships among variables and the battery discharge behavior. The goal is to predict the driving time required to go from an initial state of charge to a final one.

Table 5: Description of dataset features

Features	Units	Min	Max	Mean	Std*
Motor_energy	kWh	0,5	74,81	7,60	6,91
Battery_energy	kWh	0,9	762,25	76,86	59,92
Used_soc	%	5	76	5,7	7,03
Driv_distance	km	0,9	106,09	15,25	12,28
Start_soc	%	15	100	77,26	17,85
Start_avg_temp	°C	-6	36,24	21	6,28
speed	km/h	1	93	44,87	15,97

*Std is short for Standard deviation as described by Ji et al., 2024 [23], which is a measure of how spread-out numbers are in the data set.

Table 6: Symbols and definitions

Symbol	Description	Units
μ_j	The centroid of cluster	Same as input features
$\mathbf{C}$	The partition of observations	-
y_{predi}	The final prediction	s
$f_t(\mathbf{x}_i)$	The output of the tree	s
α_t	The associated weight	-
T	The total number of trees	-
ϵ_t	The error term for each data point	s
$g(t)$	The trend of the series	s
$s(t)$	The seasonal effects	s
$h(t)$	Model's holidays or special events	s

3.2. Data analysis

Before conducting the main analysis, the dataset undergoes a preprocessing phase consisting of the following key steps to ensure data quality and reliability.

- a. Detection and correction of anomalies: A preliminary analysis was performed on the raw dataset to identify duplicates, missing values, and transmission errors. SQL (Structured Query Language) was used to automatically detect duplicate entries and missing data. A manual line-by-line review was then conducted to ensure the dataset accurately reflected real-world conditions observed in the field. Some inconsistencies were identified, for example, when the charging status indicated that the vehicle was charging, while the SoC was decreasing and the vehicle speed was not zero. After examining the surrounding data points, these cases were interpreted as transmission errors, and the affected rows were removed from the dataset.
- b. Removal of missing values: Before proceeding, the dataset is sorted chronologically. The data is then segmented to retain only driving phases, excluding charging periods. No interpolation methods are applied; rows containing missing values are simply

removed to ensure that only real and complete data is used for the analysis.

- c. Feature engineering: Once the data is cleaned, relevant features are extracted. Since power is already provided in the dataset, the analysis focuses on computing battery energy and current consumption rate for each driving segment. These variables provide insights into the vehicle's energy behavior.
- d. Normalization: To ensure no variable dominates the learning process, Min-Max normalization is applied, as described by S. Zhao et al., 2025 [24]. This transforms all features to a comparable scale, which is particularly important for models like XGBoost that are sensitive to feature magnitude differences, according to Zhang et al., 2024 [25].

3.3. K- means clustering

To better characterize driving behaviors in the dataset, an unsupervised clustering analysis was conducted using the K-means algorithm, which seeks to partition the n observations into K clusters by minimizing the within-cluster sum of squared distances, as described by Hu et al., 2023 and Prasad et al., 2023 [26, 27], as described by formula (3).

$$\min \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - \mu_j\|^2, \quad (3)$$

where $\mathbf{C}=\{C_1, \dots, C_k\}$ denotes the partition of observations and μ_j is the centroid of cluster C_j .

The K-means algorithm was set to extract five distinct driving clusters. The number of clusters was determined using the elbow method, as proposed by R. K. Gupta et al., 2025 and F. Liu & Deng, 2021 [28, 29], which provides a good balance between model complexity and cluster separation.

The resulting five driving clusters (0 to 4), illustrated in figure 5, can be described as follows:

- Cluster 0 (blue) - Normal driving: low current and low variation, with moderate speed.
- Cluster 1 (yellow) - Fast driving: high speeds and notable variation in battery current.
- Cluster 2 (green) - Aggressive driving: high current and rapid changes, typically associated with sharp acceleration phases.

- Cluster 3 (red) - Unstable driving: erratic speed and rapid current changes, potentially due to dense urban traffic or frequent stops.
- Cluster 4 (pink) - Eco-driving: moderate to low speed, low current, and minimal fluctuations indicating energy-efficient behavior.

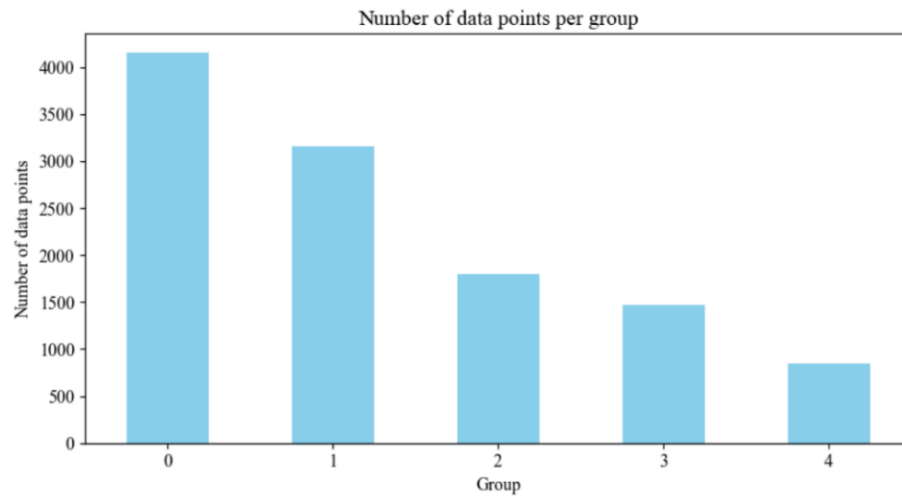


Figure 5: Number of data points per group

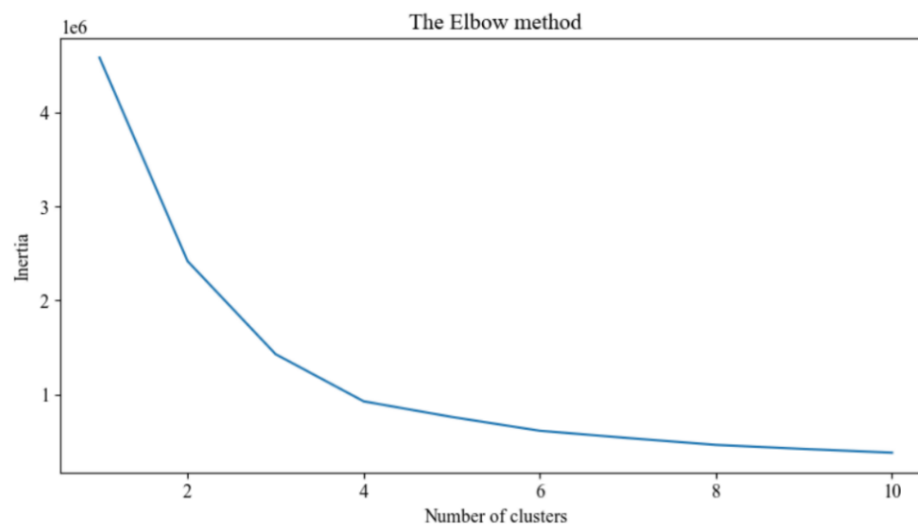


Figure 6: Elbow method for determining the optimal number of clusters

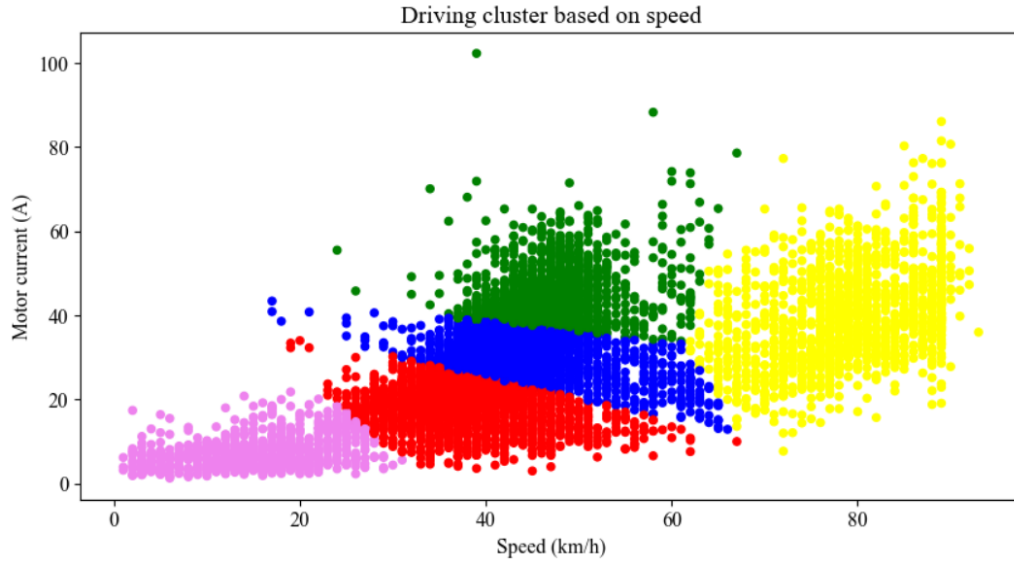


Figure 7: The five cluster group

The cluster assigned to each observation was included as a categorical variable (named group) in the predictive models to provide additional behavioral context. For sequential models such as LSTMs, the group variable was incorporated as a distinct attribute to capture driving style dynamics over time. For tree-based models like Random Forest, LightGBM, and XGBoost, group was used as an additional feature, leveraged according to each model's specifics. Finally, in the Prophet model, the variable was added as an external regressor, helping to capture variations related to driving behaviors. This differentiated integration ensures optimal use of the group variable within each modeling framework.

Figure 6 presents the elbow curve used to determine the optimal number of clusters, and the distribution of data across the five identified driving patterns.

Figure 5 and figure 7 presents the distribution of data across the five identified driving patterns. Cluster 0, representing normal driving, contains the largest number of data points.

3.4. RDT predictions models with justifications

Selecting an appropriate ML model is a critical, stand-alone phase that underpins accurate and reliable predictions and positions this work within the existing body of research. According to most studies on electric vehicles, the models adopted in this work rank among the most

frequently used and consistently demonstrate strong performance in the specialized literature, as reported by R. K. Gupta et al., 2025, Prasad et al., 2023, and L. Zhao et al., 2020 [8, 27, 29]. Therefore, in this study, five models namely, XGBoost, LightGBM, RF, LSTM and Prophet, presented in section 4.1 to 4.5 respectively, were compared. Also, the models were trained in a Python environment using a random partitioning of the data to ensure a balanced distribution between the training and test sets.

3.4.1. XGBoost

Extreme Gradient Boosting (XGBoost) is a decision tree-based boosting model widely used for regression and classification tasks due to its ability to capture complex relationships between variables. It works by constructing successive decision trees, where each new tree corrects the errors of the previous ones, thereby reducing model bias and variance. Its efficiency relies on an iterative optimization approach and effective resource management, allowing it to process large datasets quickly. One of XGBoost main advantages is its fast execution speed, achieved through parallelized training and compressed memory processing. Additionally, it incorporates regularization mechanisms (L1 and L2) to prevent overfitting and improve model generalization. It also handles missing values automatically, simplifying data preprocessing. Its flexibility allows fine-tuning of various hyperparameters, such as tree depth, learning rate, and the number of estimators, making it a preferred choice for many applications, as highlighted by M. Gupta et al., 2023 and Sun et al., 2025 [30, 31]. This model is particularly well suited for the project's dataset due to its ability to efficiently handle large volumes of data, automatically process missing values, and adapt to the complexity of variable interactions.

The boosting principle is based on aggregating predictions from multiple trees, as described by equation (4):

$$y_{predi} = \sum_{t=1}^T \alpha_t f_t(x_i), \quad (4)$$

where y_{predi} is the final prediction, $f_t(x_i)$ represents the output of tree t , α_t is the associated weight, and T is the total number of trees.

In this study, the XGBoost model was trained on the dataset by tuning the hyperparameters to optimize prediction accuracy.

3.4.2. LightGBM

Light Gradient Boosting Machine (LightGBM) is a decision tree-based boosting algorithm designed to be fast and efficient in handling large datasets. It uses a leaf-wise tree growth approach, unlike XGBoost, which employs a level-wise approach. LightGBM leaf-wise method focuses on the most informative leaves for each tree, which can result in better accuracy with fewer trees, while reducing training time, as reported by H. Liu et al., 2020 [32]. Although XGBoost and LightGBM share similar principles in gradient optimization and loss function reduction, as noted by Polat et al., 2025 [33], the difference in tree growth methods can impact performance depending on the specifics of the data and the problem being addressed. LightGBM is particularly effective on large datasets and is designed to optimize memory usage while maintaining good model generalization. In the context of this project, which involves large-scale driving data with high variability, LightGBM leaf-wise strategy enables faster training and better ability to handle very large volumes of data efficiently.

3.4.3. RF

Random Forest is an ensemble algorithm based on bagging multiple decision trees, widely used for regression and classification tasks due to its robustness against overfitting and its ability to model complex nonlinear relationships. It builds T independent decision trees, each trained on a bootstrap sample of the dataset and, at each split, selecting a random subset of features. The final prediction is obtained by aggregating the individual tree predictions, typically by majority vote for classification or averaging for regression, as reported by Acharyaviriya et al., 2023 [10]. This approach reduces model variance without substantially increasing bias, thanks to the randomness introduced in both sampling observations and selecting features. Moreover, Random Forest provides an internal measure of feature importance, aiding interpretability and feature selection.

The aggregated regression prediction is given by:

$$\hat{y}_i = \frac{1}{T} \sum_{t=1}^T f_t(x_i), \quad (5)$$

where $f_t(x_i)$ is the prediction of tree t for observation x_i , and T is the total number of trees.

3.4.4. LSTM

LSTM (Long Short-Term Memory) is a deep learning model designed for processing and forecasting complex time series data. Developed as part of the recurrent neural network (RNN) family, it addresses the traditional limitations of RNNs in capturing long-term dependencies by incorporating memory cells and gated mechanisms (input, output, and forget gates), as described by Koohfar et al., 2023 [34]. This structure enables LSTM to retain relevant information over long time intervals, making it particularly suitable for analyzing chronological data with nonlinear or long-range dependencies.

LSTM is widely applied in contexts where the influence of past events on future outcomes is not immediate, such as in energy consumption forecasting, user behavior modeling, or the evolution of battery charge states. Unlike traditional models that require explicit decomposition into trend or seasonal components, LSTM learns temporal patterns automatically, as highlighted by Khan, 2020 [35]. In the context of maintaining the range of EV, LSTM can model the evolution of the SoC based on the provided data, such as driving styles and driving conditions. This capability enhances the accuracy of battery behavior predictions under dynamic conditions, thereby enabling more effective range management.

3.4.5. Prophet

Prophet is a model developed by Facebook, designed for time series forecasting, as reported by Yapıcı et al., 2025 [6]. It is particularly suited for analyzing complex chronological data, such as those observed in contexts where seasonal trends, the effects of holidays, and other external events influence the data. One of the main advantages of Prophet is its ability to model time series with nonlinear behaviors and to effectively capture long-term variations. This model is robust and flexible, making it effective for high-frequency time series with

numerous irregularities, as reported by Baranowski et al., 2021 [36]. The particularity of Prophet lies in its method of decomposing time series into three main components: trend, seasonality, and external factors. This allows the model to adapt to rapid changes in data while maintaining good long-term forecasting. The model also uses timestamps to better capture the chronological evolution of events, making it particularly suitable for applications such as demand forecasting, inventory management, or analyzing human behaviors in various contexts, as reported by Serrano et al., 2025 [37]. In the case of electric vehicles (EVs), this ability to capture chronological changes allows for the analysis of driving habits, and variations in battery state of charge, while accounting for weather conditions and specific events. This enables better estimation of battery behavior in dynamic contexts. The equation for Prophet can be expressed as follows:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t, \quad (6)$$

where $y(t)$ is the value of the time series at time t , $g(t)$ represents the trend of the series, $s(t)$ captures the seasonal effects, $h(t)$ model's holidays or special events, ϵ_t is the error term for each data point.

3.5. Cross-validation

Cross-validation is a model evaluation technique in which the dataset is partitioned into k subsets (folds), and training/testing is repeated l times, each fold serving once as the test set while the other $k-1$ folds are used for training. This procedure reduces the variance of performance estimates and guards against overfitting by leveraging all data for evaluation, as demonstrated by Oyucu et al., 2024 and L. Zhao et al., 2020 [8, 38]. The five folds used in this procedure are presented in figure 8, which illustrates how the dataset is divided and rotated across iterations.

A 5-fold cross-validation was performed on 15000 samples. Each fold thus contains 3000 rows: the model is trained on 12000 points and tested on 3000 in each iteration. Selecting $k=5$ is a common trade-off between estimation reliability (lower bias) and computational cost (moderate variance).

The same data splits were used across all models to ensure that performance comparisons are based on identical training and testing sets.

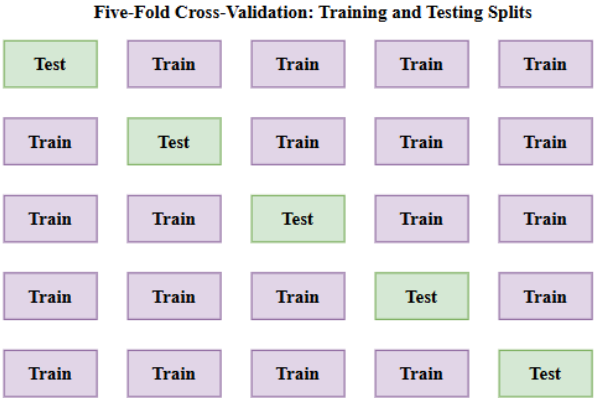


Figure 8: Data splitting into training and test sets for cross-validation

3.6. Hyperparameter selection

The selection of hyperparameters is a crucial step in building an effective model, as it directly influences its ability to capture underlying data patterns while avoiding overfitting.

To this end, a systematic strategy based on grid search combined with cross-validation was adopted to determine the optimal values of the hyperparameters.

A grid of values was defined for each hyperparameter (notably the learning rate, batch size, and regularization coefficients), and all possible combinations were tested.

Each combination was evaluated using cross-validation, providing a reliable estimate of the model's performance and generalization capability.

The selected configuration corresponds to the one that achieved the best average results across all validation folds. The test set was kept exclusively for the final evaluation, ensuring an unbiased assessment of performance.

The hyperparameter values that led to the best performance were retained and are presented in table 7.

Table 7: Parameters values used in the machine learning models

Models	Parameters
LightGBM	n_estimators: 700, Learning_rate: 0.05, Max_depth: 6, Max_features: 8, Subsample: 1.0, Early_stopping_rounds=10
XGBoost	n_estimators: 700, Learning_rate: 0.05, Max_depth: 6, Max_features: 8 subsample: 1.0, Early_stopping_rounds=10
Prophet	seasonality_mode=multiplicative, Changepoint_prior_scale=0.5, changepoint_prior_scale=0.5, Seasonality=daily, weekly
LSTM	epochs=50, Model_type=sequential, Lstmunits=50, Returnsequence=true Dropout_rate=0.2, Optimizer=adam, Early_stopping=true
RF	n_estimators=700, max_depth=6, max_features=8, bootstrap=True, random_state=42, n_jobs=-1

3.7. Performances metrics

To evaluate the performance of the predictive models developed for RDT estimation, several performance metrics are used. These metrics measure the discrepancy between the predicted values and the actual values, thereby assessing the accuracy and robustness of the models. Three main indicators are considered in this study: the Mean Absolute Error (section 3.7.1), the Root Mean Square Error (section 3.7.2), and the coefficient of determination (section 3.7.3).

3.7.1. Mean Absolute Error (MAE)

MAE is a widely used performance metric for evaluating the accuracy of a regression model. It measures the average difference between predicted and actual values by taking the absolute value of each error, preventing positive and negative errors from canceling out. Unlike RMSE, MAE treats all errors equally without giving more weight to larger deviations. A lower MAE value indicates higher model accuracy, while a higher value suggests less reliable predictions. This metric is particularly useful when the goal is to minimize the average

error without being overly influenced by outliers, as shown by Oyucu et al., 2024 [38]. The MAE is defined by the following equation:

$$MAE = \frac{\sum_{i=1}^n |y_{réel} - y_{hat_final}|}{n}, \quad (7)$$

where $y_{réel}$ represents the actual values, y_{hat_final} the predicted values of the model.

3.7.2. Root Mean Squared Error (RMSE)

RMSE is a commonly used performance metric for evaluating the accuracy of a regression model. It measures the average difference between predicted and actual values, giving greater weight to larger errors due to the squared differences. A lower RMSE value indicates that the model predictions are closer to the observed values, while a higher value suggests greater error dispersion. This metric is particularly useful for comparing different models or optimizing an model to improve its accuracy, as demonstrated by Achariyaviriya et al., 2023 [10]. The RMSE is defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{réel} - y_{hat_final})^2}{n}}, \quad (8)$$

where $y_{réel}$ represents the actual values, y_{hat_final} the predicted values of the model.

3.7.3. Coefficient of Determination (R^2)

R^2 is a commonly used performance metric for evaluating the quality of a regression model. It measures the proportion of the variance in the dependent variable that is explained by the independent variables in the model. Its value ranges from 0 to 1, where a value close to 1 indicates that the model explains a large part of the data variability, while a value close to 0 suggests that it fails to capture this variability effectively, as noted by Adhya et al., 2022 [39]. The R^2 is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{réel} - y_{hat_final})^2}{\sum_{i=1}^n (y_{réel} - y_{moyenne})^2}, \quad (9)$$

where $y_{réel}$ represents the actual values, $y_{hat\ final}$ the predicted values of the model and $y_{moyenne}$ the mean of the observed values.

3.8. Statistical tests

To rigorously compare the performance of the different models, we employed non-parametric statistical tests suitable for repeated measure evaluations. All models were evaluated on the same five cross-validation folds, which places the comparison in a repeated-measures setting where each fold acts as a paired observation across models. Because RMSE, MAE and R^2 values obtained across folds do not satisfy the normality assumption required for parametric tests such as repeated-measures ANOVA, a non-parametric procedure was required. Since model performance distributions cannot be assumed to be normal and all models are assessed on the same experimental repetitions, the Friedman test was applied to compare the performance of the prediction models (XGBoost, LightGBM, Prophet, LSTM and RF) across all cross-validation scores. This nonparametric test is suitable for comparing multiple models evaluated on the same validation folds, as it does not rely on the assumption of normally distributed. A significance threshold of 0.05 was adopted.

$$\chi_F^2 = \frac{12N}{k(k+1)} \left(\sum_{j=1}^k \bar{R}_j^2 \right) - 3N(k+1), \quad (10)$$

where N denotes the number of cross-validation folds, K represents the number of models included in the comparison, $\bar{R}_j$ indicates the average rank obtained by model j across all folds. When the Friedman test indicates a significant overall effect, a post-hoc analysis is carried out using the Nemenyi test. This test performs pairwise comparisons based on average ranks while appropriately controlling for multiple comparisons. It allows us to identify which specific pairs of models differ significantly in performance. This test relies on the Critical Difference (CD), defined as:

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}}, \quad (11)$$

where q_α is the critical value obtained from the studentized range distribution for significance level α , k corresponds to the number of models, N denotes the number of cross-validation folds. This two-step statistical protocol is widely adopted in comparative studies of machine learning models, as illustrated in Viana dos Santos Santana et al., 2021 [40].

The Friedman and Nemenyi tests were chosen because they are well suited for comparing multiple models evaluated on the same datasets, even when performance measures do not follow a normal distribution. These tests therefore provide a robust statistical analysis in this context, whereas parametric methods remain appropriate when their assumptions, such as data normality, are met.

CHAPITRE 4

4. RESULTS AND DISCUSSIONS

This section presents and discusses the main results obtained during the experimental phase of this study. It is structured into three parts. First, the model's performance is evaluated using cross-validation to assess its generalization ability (section 4.1). Then, feature importance is analyzed using the SHAP model, to identify the most influential factors in the prediction process (section 4.2). The statistical comparison of the algorithms is then performed using the Friedman test (section 4.3), followed by post-hoc analysis with the Nemenyi test (section 4.4). Finally, a driving experiment is carried out until full battery discharge to evaluate the model's ability to estimate the total driving duration (section 4.5).

4.1. Cross-validation

The results obtained demonstrate the generalization of the XGBoost, LightGBM, LSTM and Prophet models in estimating the remaining driving time of a battery based on driving conditions and environmental characteristics.

As seen in table 8 and figure 13, several performance metrics were computed across five cross-validation folds to evaluate the accuracy and robustness of the models in predicting the driving time between two states of charge of the battery.

Among the tested models, Random Forest (RF) demonstrated very good performance with a mean RMSE of 0.0215, MAE of 0.0125, and a determination coefficient (R^2) of 0.9430. However, this model comes with a relatively long runtime (306.22 seconds) and higher memory usage (26.21 MB).

The LSTM model achieved the highest R^2 score (0.9517), reflecting excellent predictive ability. However, this came at the cost of significant computational demand, with an average runtime of 516.47 seconds and memory usage reaching 49.23 MB. While capable of learning temporal dependencies in time series data, its heaviness may limit real-time applications.

XGBoost and LightGBM performed well, with R^2 scores of 0.9480 and 0.9446, respectively, while maintaining low execution times (26 seconds) and modest memory usage (17 MB). Their

effectiveness demonstrates the ability of boosting models to handle moderate nonlinear relationships, offering a good balance between accuracy and execution time.

The Prophet model stands out for its ability to minimize absolute errors, with a mean RMSE of 0.0272 and MAE of 0.0182, making it one of the top performers on these metrics. Although its R^2 (0.9113) is slightly lower than other models, it remains competitive in prediction quality. Prophet also benefits from a fast runtime (78.64 seconds) and low memory usage (14.83 MB), making it suitable for resource-constrained environments. Its capacity to model temporal trends and seasonal effects, while maintaining low errors, enhances its relevance for time series forecasting despite its simpler structure limiting complex interaction modeling.

In summary, while all models demonstrated satisfactory performance, XGBoost and LightGBM offer a good balance between accuracy and execution time. Random Forest and LSTM stand out for high accuracy but come with higher computational costs. Finally, Prophet, although lightweight, remains a valid option in specific contexts.

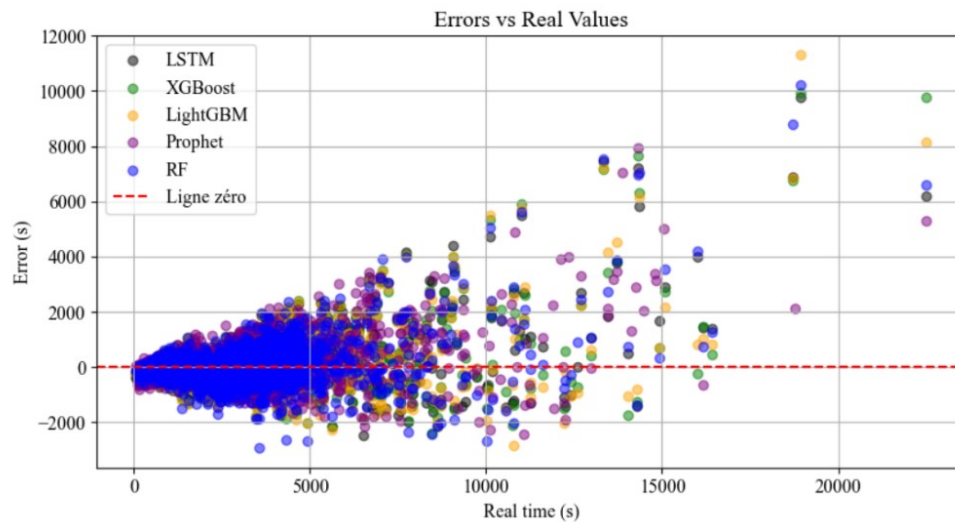


Figure 9: Error dynamics during real-time evaluation

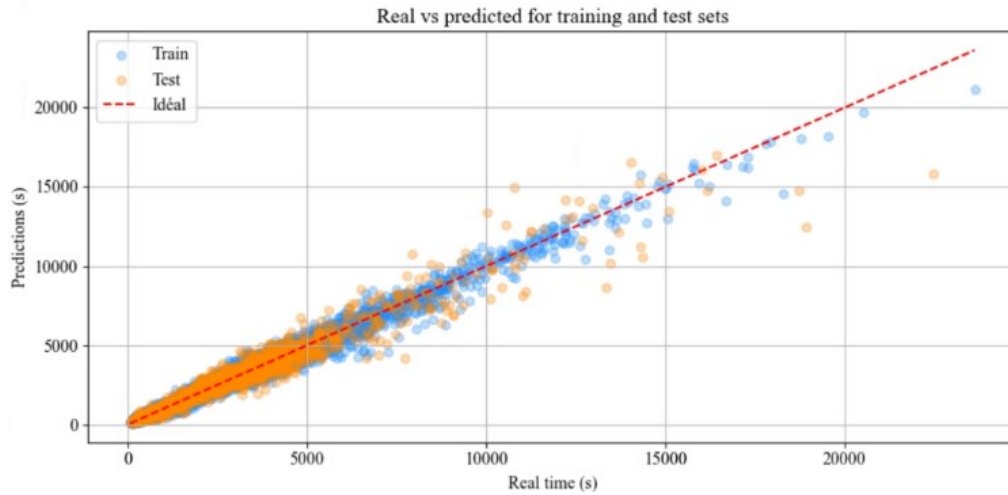


Figure 10: Predicted vs actual values for training and test sets

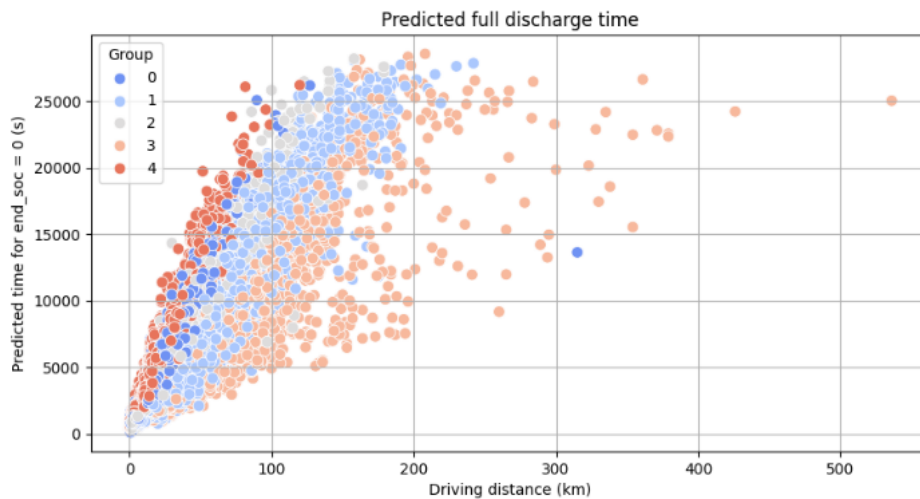


Figure 11: Predicted time vs distance for full battery discharge

The analysis of prediction errors for the time between two state-of-charge levels is shown in figure 9, where a strong concentration of errors closes to zero is observed for short durations, while the dispersion increases for higher actual times. The figure 14 complements this view by presenting separate scatter plots for each model, allowing identification of model-specific variations in the errors. In both figures, the error distribution shows that the models remain generally balanced without systematic bias but encounter a small difficulty in accurately predicting longer durations.

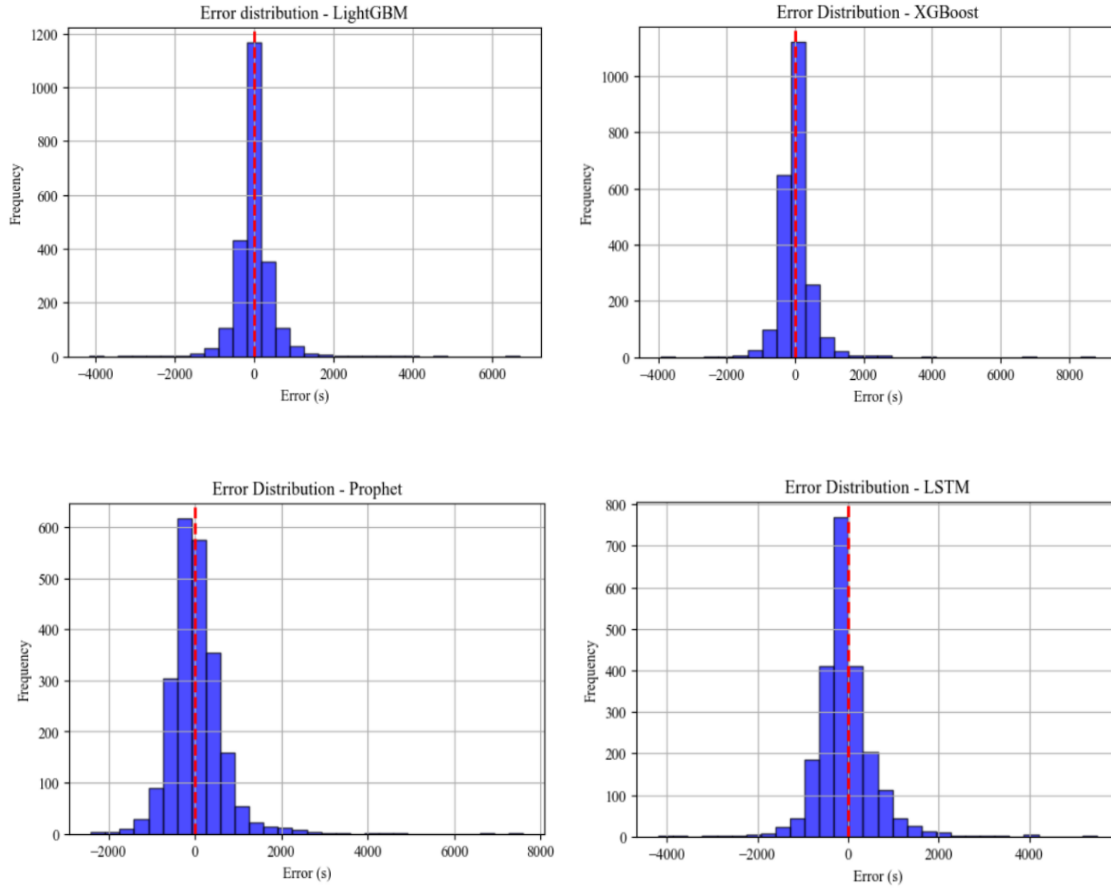


Figure 12: Error dynamics of each model during real-time evaluation

For better observation, a comparison curve between the actual and predicted values was generated, as shown in figure 10. This curve highlights the model's accuracy and its ability to follow data trends, thereby confirming the effectiveness of the approaches used. Moreover, the closeness of the training (blue) and testing (orange) points to the ideal line (red dashed) demonstrates similar performance on both datasets, indicating that there is no overfitting. The model therefore generalizes well to new data. The impact of weather conditions, especially extreme temperatures, was a significant factor in improving the accuracy of predictions. The integration of these data into the boosting models allowed for better handling of these discrepancies, providing more reliable forecasts in extreme conditions. Driving habits also influenced the prediction results. Drivers with more stable driving styles allowed the model to predict the remaining battery discharge time more accurately. In contrast, drivers with erratic

driving behaviors, including rapid accelerations and frequent braking, led to greater variability in prediction results. This phenomenon was particularly well captured by the boosting models, which were able to adjust predictions based on different energy consumption profiles. The curves generated from the actual and predicted values showed that, while most predictions were close to the actual values, some discrepancies remain, particularly in situations where driving conditions change rapidly. These observations highlight the importance of further improving the models by considering additional variables or refining the models to better capture these rapid dynamics.

In summary, although XGBoost and LightGBM demonstrated slightly superior performance compared to Prophet, the latter remains a relevant choice in specific situations where modeling temporal trends is crucial. The integration of environmental variables and driving habits into the models clearly improved predictions, emphasizing the need for a multimodal approach in predicting the remaining driving time under real-world conditions.

Table 8: Results of the cross-validation experiment

Method	Evaluation function	1 st Fold	2 nd Fold	3 rd Fold	4 th Fold	5 th Fold	Mean	Runtime (s)	RAM (MB)
XGBoost	RMSE	0,2534	0,2185	0,2178	0,2324	0,2163	0,2277	26,51	17,83
	MAE	0,1434	0,1318	0,1334	0,1418	0,1386	0,1378		
	R^2	0,9401	0,9529	0,9471	0,9501	0,9499	0,9480		
LightGBM	RMSE	0,2604	0,2297	0,2313	0,2333	0,2203	0,2350	26,29	17,41
	MAE	0,1432	0,1321	0,1352	0,1413	0,1377	0,1379		
	R^2	0,9368	0,9479	0,9403	0,9497	0,9481	0,9446		
Prophet	RMSE	0,0270	0,0293	0,0260	0,0245	0,0292	0,0272	78,64	14,83
	MAE	0,0183	0,0183	0,0176	0,0177	0,0190	0,0182		
	R^2	0,9189	0,9009	0,9078	0,9219	0,9071	0,9113		
LSTM	RMSE	0,2214	0,2129	0,2127	0,2368	0,2123	0,2192	516,47	49,23
	MAE	0,1492	0,1359	0,1351	0,1518	0,1386	0,1421		
	R^2	0,9551	0,9554	0,9482	0,9493	0,9509	0,9517		
RF	RMSE	0,0243	0,0223	0,0207	0,0208	0,0195	0,0215	306,22	26,21
	MAE	0,0128	0,0136	0,0118	0,0123	0,0122	0,0125		
	R^2	0,9326	0,9397	0,9415	0,9512	0,9500	0,9430		

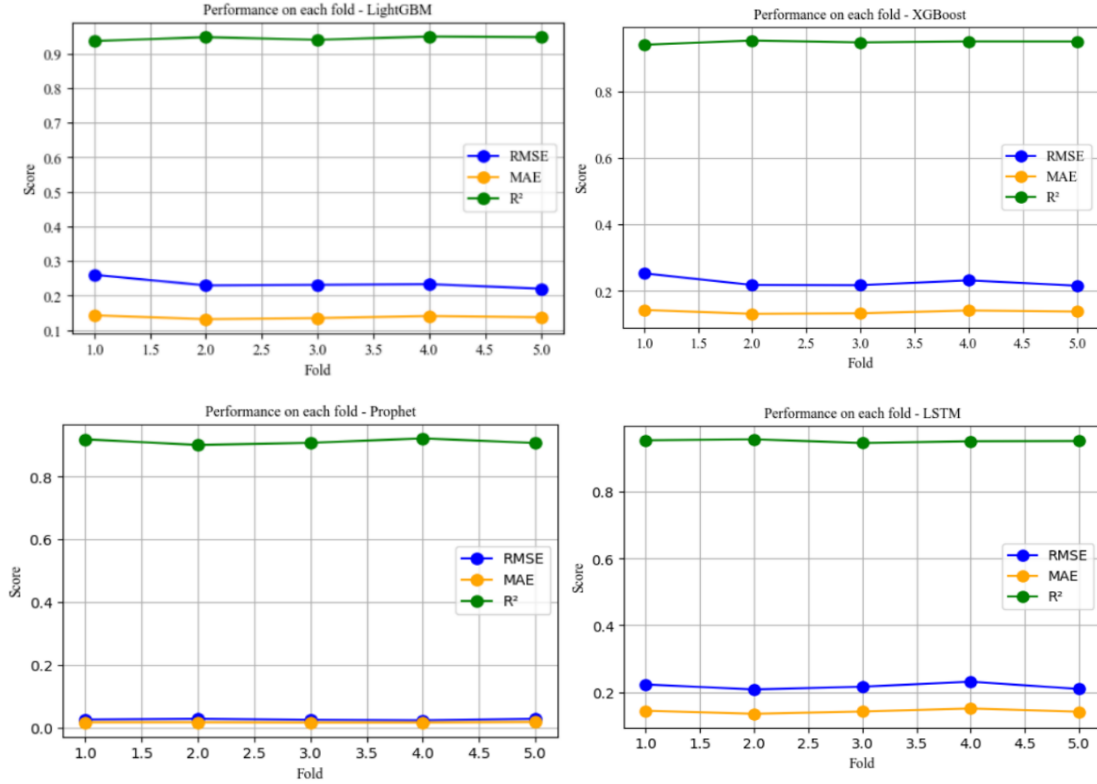


Figure 13: Cross-validation performance on each fold.

The distribution of prediction errors for each model is shown on the histogram in figure 12. For all methods, the errors are mostly centered around zero, reflecting good overall accuracy. The histograms display relatively symmetrical shapes, with a high frequency of small errors and few extreme values. This indicates that the models do not exhibit significant systematic bias and that most predictions remain close to the actual values. These results confirm the stability of the models in predicting the time between two state-of-charge levels.

Overall, the distributions are narrow, indicating limited variability in the prediction errors. The concentration around zero suggests strong generalization capabilities on the test data. This visualization offers a concise overview of the consistency in predictive performance across the different models.

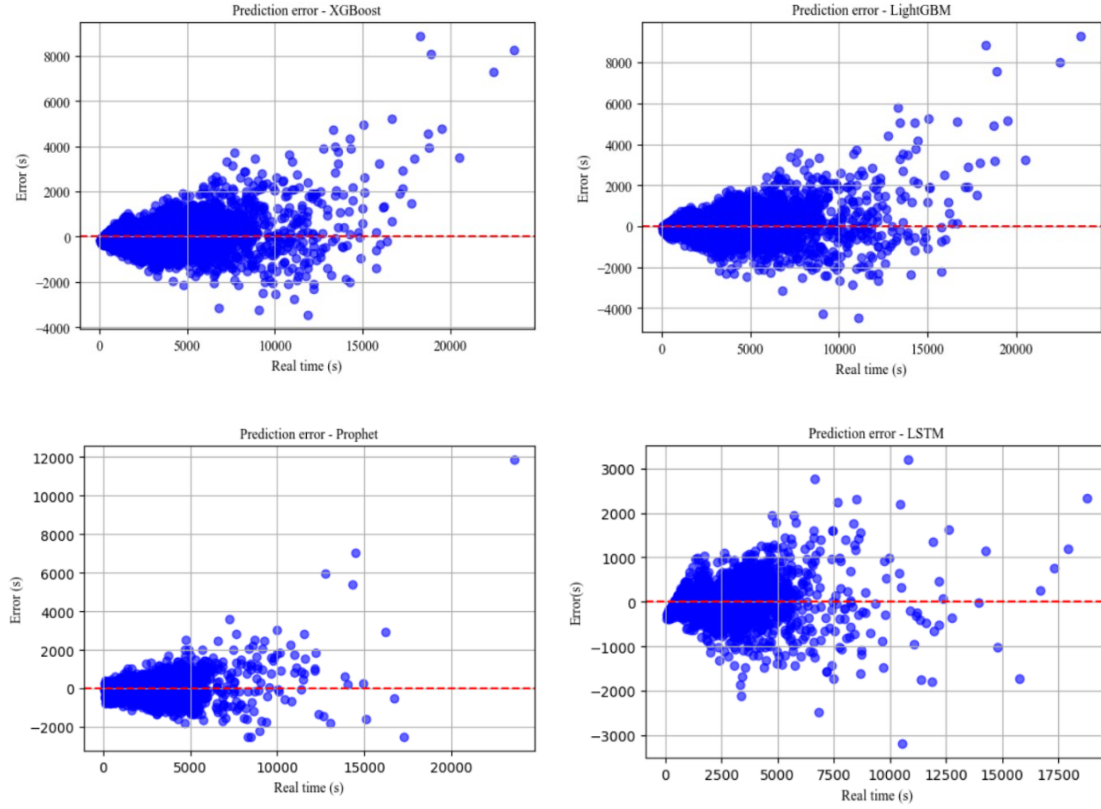


Figure 14: Cross-validation real-time tracking of prediction error

4.2. Features selection

The SHAP (Shapley Additive exPlanation) method is used to identify the importance of features in a predictive model. Gebreyesus et al., 2023 and F. Liu & Deng, 2021 [28, 41] describe this approach as being based on the Shapley value theory from cooperative game theory, where a specific contribution is assigned to each variable. The calculation relies on the average differences between the actual and predicted values while considering all possible permutations of the variables.

Thus, SHAP quantifies the individual influence of each feature on the model output, ensuring robust interpretability of the results. As seen in figure 15, features with a strong impact on the target feature appear at the top, while those with low correlation are positioned bottom. Furthermore, the color coding helps interpret the significant of each feature, with red and blue color indicating a high and low correlation respectively. **Battery_energy**, **driv_distance**, **motor_energy** and group are among the most significant features on the vehicle remaining

vehicle charge time. Also, features, like **end_soc**, and **max_avg_temp** have a moderate effect but remain significant, as they reflect environmental and geographical conditions that influence the vehicle energy consumption. Generally, all the features show a clear relationship with the remaining vehicle driving time prediction, each having a distinct impact that varies in intensity, as revealed by the SHAP analysis.

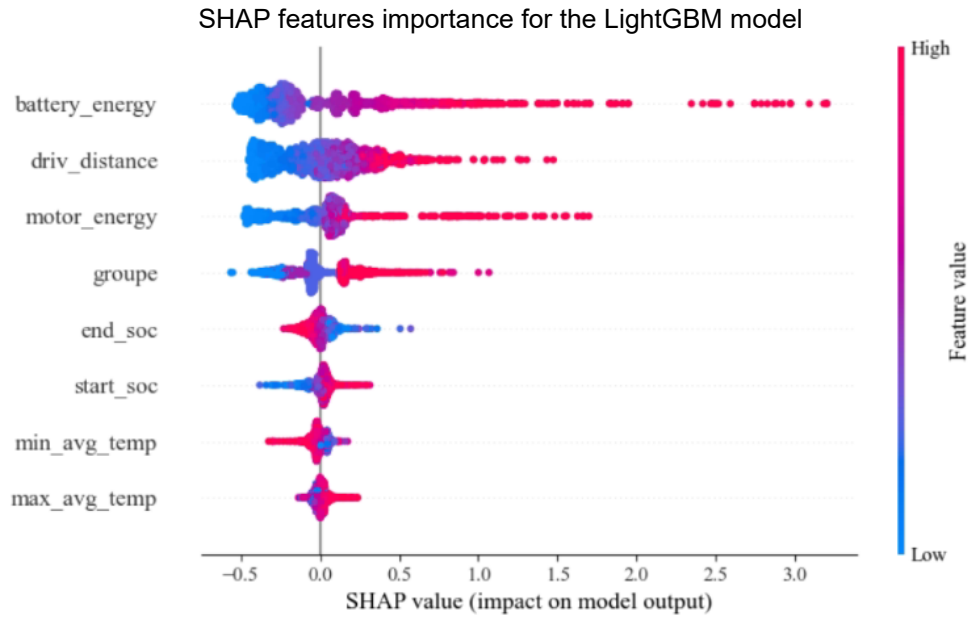


Figure 15: SHAP Summary Plot of Feature Impact on Model Output

4.3. Friedman test

The Friedman test was applied to the three metrics (R^2 , RMSE, and MAE) to determine whether significant differences existed between the models. The results are presented in table 9.

Table 9: Friedman test results

Metric	Friedman statistic	p-value
R^2	13,60	0,0087
RMSE	18,08	0,0012
MAE	17,66	0,0014

The Friedman statistics are 13.60 for R^2 ($p = 0.0087$), 18.08 for RMSE ($p = 0.0012$), and 17.66 for MAE ($p = 0.0014$). The statistic measures the deviation of the algorithm's average ranks, and the associated p-value, here all p-values are below 0.05, indicating clear overall differences

across the algorithms.

The strongest effects appear for RMSE and MAE, whose very low p-values highlight substantial variability in model performance. These results show that the Friedman test rejects the null hypothesis of equal model performance, thereby motivating the use of a post-hoc comparison, such as the Nemenyi test, to determine which model pairs differ significantly and to identify the best-performing models statistically.

4.4. Nemenyi test

To examine in detail the differences in model performance, the results of the Nemenyi test are presented in the following three subsections: section 4.4.1 Nemenyi for R^2 , section 4.4.2, Nemenyi for RMSE and section 4.4.3 Nemenyi for MAE. Each subsection analyzes the pairwise p-values and highlights statistically significant differences when present.

4.4.1. Nemenyi test for R^2

The Nemenyi test results for R^2 are presented in figure 16. Prophet is significantly different from several other models. Comparisons between XGBoost and Prophet with a p-value around 0.04088 and LSTM and Prophet with a p-value around 0.00607 highlight marked differences. In contrast, XGBoost, LightGBM, LSTM, and RF show high p-values among themselves, indicating statistically similar performance for this metric. These results suggest that Prophet performs significantly worse while the other models can be considered equivalent in terms of R^2 .

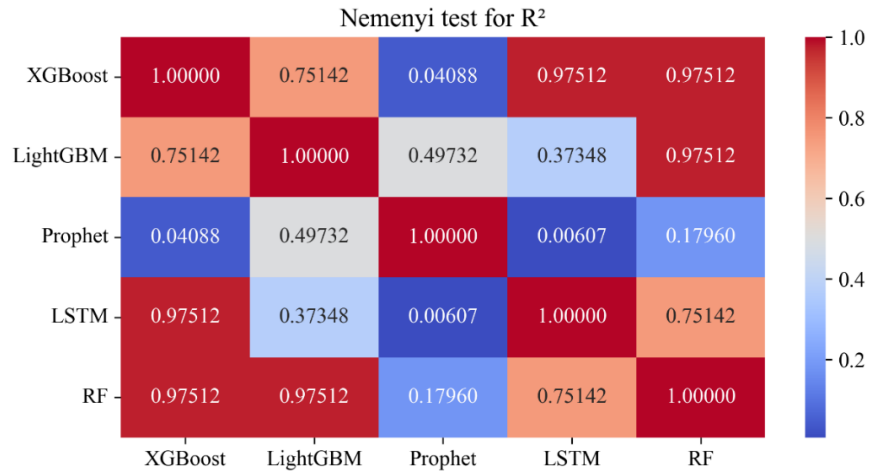


Figure 16: Nemenyi test for R square

Red indicates high p-values, meaning no significant difference, whereas pure blue indicates low p-values, revealing statistically significant differences between the models.

4.4.2. Nemenyi test for RMSE

The Nemenyi test results for RMSE are presented in figure 17. LightGBM and RF stand out clearly from the other models with very low p-values in their comparisons, notably between RF and LightGBM ($p = 0.00136$) and between RF and XGBoost ($p = 0.04088$). These results indicate that RF performs significantly differently, often better, compared to several other models. In contrast, comparisons between XGBoost, LightGBM, Prophet, and LSTM show higher p-values, suggesting no significant differences among these models for this metric.

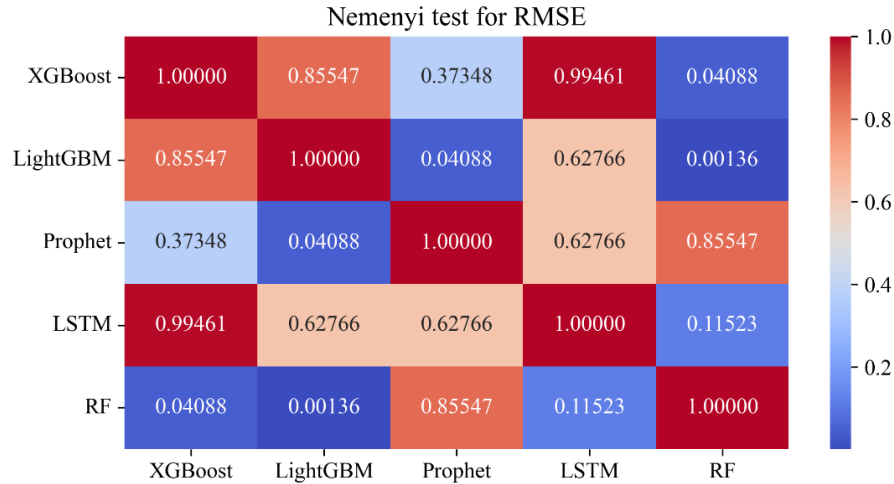


Figure 17: Nemenyi test for RMSE

4.4.3. Nemenyi test for MAE

The Nemenyi test results for MAE are presented in figure 18. It can be observed that RF and LSTM clearly stand out from the other models, with very low p-values in their comparisons, notably between RF and LSTM ($p = 0.00201$) as well as between RF and XGBoost ($p = 0.05394$). These results indicate that RF and LSTM perform statistically differently, often better, compared to several other algorithms. In contrast, comparisons between XGBoost, LightGBM, and Prophet show higher p-values, suggesting generally similar performance among these models for this metric.

The combined analysis of the Nemenyi post-hoc tests for R^2 , RMSE, and MAE shows that some models differ significantly while others exhibit statistically similar performance. Prophet appears as a less effective model in terms of R^2 , whereas RF performs favorably on the RMSE and MAE metrics, often accompanied by LSTM for MAE. XGBoost and LightGBM display performance generally comparable to the other algorithms across all metrics. These observations will be complemented by considering execution time in the overall assessment.

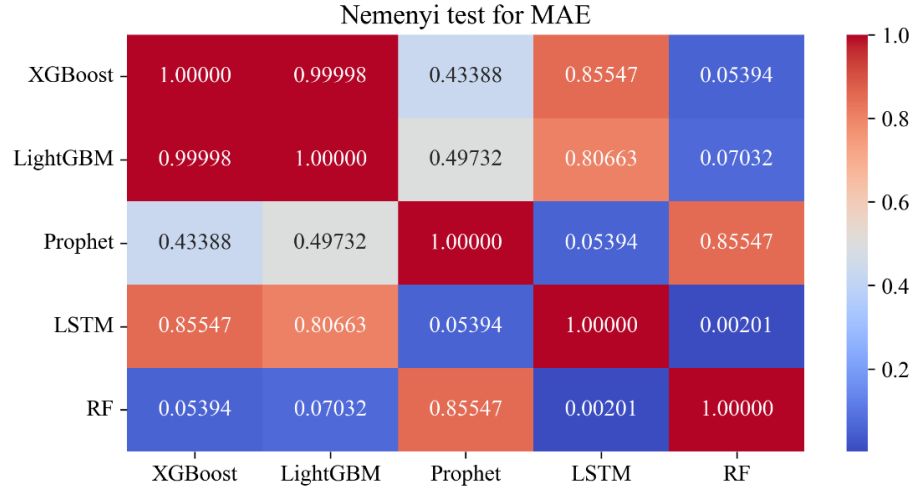


Figure 18: Nemenyi test for MAE

4.5. Prediction of driving time for end SoC=0

The prediction of driving time until a final SoC of 0% is carried out using supervised machine learning methods, including XGBoost, LightGBM, Random Forest and Prophet. These models perform a progressive extrapolation based on partially observed discharge trajectories, learning to estimate the remaining driving time without relying on fixed rules or analytical models. This approach leverages patterns learned from historical data to generalize across different driving conditions. The dataset used for training is structured and summarized in table 5 outlining the features considered. Model performance is evaluated using standard metrics such as RMSE, MAE and R^2 , ensuring a reliable comparison between approaches.

The performance of the different models applied to the prediction of the remaining time until full battery discharge is summarized in table 10. These results were obtained using the training set only, as the available data do not include any instances where the battery is fully discharged.

Among the models tested, LSTM shows the best performance in terms of accuracy, with an RMSE of 0.22 and an R^2 of 0.9526, suggesting a strong capacity to model temporal dynamics, although this model requires a longer execution time compared to the others. LightGBM, XGBoost, and Random Forest also yield very similar performance (RMSE = 0.23, MAE = 0.13, R^2 around 0.95), while offering shorter execution times, especially for LightGBM. Prophet, on the

other hand, produces significantly less accurate results (RMSE of 0.54 and R^2 of 0.7257), indicating that it is less well-suited to the predictive task under study.

Table 10: Performance metrics

Models	Suggested approaches		Reference articles	
	Metrics	Runtime (s)	Metrics	Articles
LightGBM	RMSE: 0,23	40,89	RMSE:1,32	[8]
	MAE: 0,13		MAE:1,29	
	R^2 :0,9515		-	
XGBoost	RMSE: 0,23	43,02	RMSE:1,35	
	MAE: 0,13		MAE:1,27	
	R^2 : 0,9498		-	
Prophet	RMSE: 0,54	125,98	-	-
	MAE: 0,46		-	
	R^2 : 0,7257		-	
LSTM	RMSE: 0,22	697,33	-	-
	MAE: 0,13		-	
	R^2 : 0,9526		-	
RF	RMSE: 0,23	324,10	RMSE: 51,98	[10]
	MAE: 0,13		-	
	R^2 : 0,9509		R^2 : 0,9261	

The relationship between predicted full discharge time and driving distance, illustrated in figure 11, highlights the variability across different driving groups. Each point represents an individual prediction, color-coded according to the group classification. The strong positive correlation observed confirms that the models capture the expected trend: the longer the driving distance, the longer the remaining time before the battery reaches a 0% state of charge. However, the dispersion of points, especially at higher distances, reflects inherent variability in driving conditions and battery usage, as well as the limitations of extrapolation when extending beyond the range of observed data. This visualization confirms the models' ability to generalize from partial discharge trajectories to provide meaningful predictions across diverse scenarios.

The variations observed in RMSE and MAE values across *figure 13* and table 10 are mainly due to differences in the data ranges used during model training and testing. Specifically, some datasets correspond to wider SoC intervals or contain higher variability in driving conditions, which tends to increase prediction error. In addition, slight differences in hyperparameter tuning

across models also contribute to these variations. These discrepancies highlight the sensitivity of certain algorithms, such as LSTM and Random Forest, to data distribution and parameter optimization.

CONCLUSION

Accurate prediction of driving time between two SoC is a key step in improving energy management for electric vehicles. In the initial phase, the prediction was performed between a starting SoC and an arrival SoC. Results show that the Prophet model achieves the lowest error in this context, although its computation time is slightly longer than that of LightGBM.

Next, cross-validation was applied to assess the robustness and generalization capabilities of the models. Prophet maintained low error, confirming its stability, while LightGBM remained competitive by offering a strong balance between accuracy and computational efficiency.

As part of the interpolation process, aimed at estimating the driving time required to reach an end SoC of 0% from a given initial SoC, the LSTM model was also tested. Although LSTM produced good predictive performance, its extremely high computation time limits its practicality in scenarios requiring fast predictions.

Considering the results of the Friedman and Nemenyi statistical tests, XGBoost, LightGBM and LSTM both show performance that is significantly different from the other models. XGBoost achieved the highest R^2 value among the tested models, indicating slightly superior fitting performance. However, LightGBM achieved nearly comparable accuracy ($R^2 = 0.9446$, $MAE = 0.1379$) while maintaining significantly lower computation time and memory usage. Therefore, when both accuracy and efficiency are considered, LightGBM represents the best trade-off between predictive performance and computational cost. This balance makes it particularly suitable for real-time or embedded energy management systems in electric vehicles.

The evaluation criteria considered in this study include predictive accuracy (R^2 , MAE, RMSE), computation time, and memory consumption, allowing for a comprehensive comparison of model performance in both accuracy-oriented and resource-constrained contexts. The comparative analysis

of model performance and resource usage highlights significant trade-offs. For instance, although LSTM achieves the lowest RMSE and MAE values, its runtime exceeds 516 seconds more than five times that of LightGBM and XGBoost making it impractical for time-sensitive tasks. Similarly, Random Forest offers competitive accuracy but suffers from long execution times and high memory usage. Prophet, while less accurate than LightGBM and XGBoost, demonstrates strong stability at the cost of higher peak memory consumption.

In summary, each model presents specific strengths and limitations. Prophet excels in accuracy for direct SoC-to-SoC predictions but requires more computational resources. LSTM is suitable for complex interpolation tasks but is hindered by its high computational cost. XGBoost achieves the best overall accuracy, while LightGBM offers the best compromise by combining strong predictive power, low error metrics, and fast computation times. This balance is crucial for effective energy management and reliable planning in electric vehicles, ensuring both precision and operational feasibility.

It should be noted that the impact of additional variables, such as road slope and traffic conditions, has not yet been evaluated in the current model. Their integration will be part of future work, with the expectation that these features could further improve the robustness and accuracy of the predictions. Other types of models will also be explored, including classical statistical approaches such as ARIMA, as well as hybrid models combining, for example, Prophet and XGBoost. The objective will be to maintain a robust balance between accuracy, interpretability, and execution time, in line with recent studies emphasizing multi-criteria evaluation in travel time prediction models.

APPENDIX

Computation Platform Details: The experiments were conducted on a VMware virtual machine running Microsoft Windows 10 Education N. The programming language used is Python 3.9.20, executed within the Jupyter Notebook 7.2.2 environment. The system is equipped with two Intel 64 Family 6 processors with a frequency of 2.99 GHz and 16 GB of RAM.

Cross-Validation: The program performs predictions on each fold during the cross-validation process. The final performance result is obtained by averaging the outcomes across all folds.

CHAPITRE 5

5. CONCLUSION GÉNÉRALE

Ce mémoire s'est intéressé à la prédiction du temps de conduite restant (RDT – *Remaining Drive Time*) dans les véhicules électriques. Une estimation fiable de ce temps constitue un enjeu majeur pour la gestion énergétique et la planification de la conduite, car elle permet d'optimiser l'autonomie et d'améliorer la fiabilité des prévisions d'utilisation du véhicule. L'objectif principal de ce travail était de comparer différentes approches de modélisation afin d'identifier la méthode la plus performante et la plus adaptée aux besoins de la gestion énergétique.

Deux types de prédiction ont été abordés : la première consiste à estimer le temps de conduite restant entre deux niveaux de charge (SoC), tandis que la seconde vise à prédire le temps de conduite jusqu'à la décharge complète du véhicule. Plusieurs modèles ont été évalués dans ce cadre, soit Prophet, LightGBM, XGBoost, Random Forest et LSTM. Les résultats ont montré que le modèle Prophet se distingue par sa stabilité et sa précision pour la prédiction directe entre deux SoC, tandis que LightGBM offre un excellent compromis entre exactitude, rapidité de calcul et efficacité computationnelle. En revanche, le modèle LSTM, bien que très précis, présente un temps d'exécution élevé (plus de 500 secondes), ce qui limite son utilisation dans les scénarios nécessitant une exécution rapide.

L'analyse comparative des performances et de la consommation de ressources a permis de dégager une hiérarchie claire :

- LSTM se montre le plus précis mais le plus coûteux en temps de calcul ;
- Prophet garantit une bonne stabilité mais requiert davantage de mémoire ;
- LightGBM s'impose comme la solution la plus équilibrée, avec une MAE de 0,1379 et un R^2 de 0,9446, tout en maintenant des temps de calcul faibles.

Ainsi, LightGBM apparaît comme la solution la plus adaptée pour la prédiction du temps de conduite restant, que ce soit entre deux niveaux de charge ou jusqu'à la décharge complète, en offrant un compromis efficace entre précision et performance computationnelle.

Au-delà de ces résultats, ce travail met en lumière les défis liés à la modélisation du temps de conduite restant dans le domaine de la mobilité électrique. Il s'agit non seulement d'améliorer la

précision prédictive, mais aussi de mieux prendre en compte la diversité des facteurs influençant le comportement du véhicule. Parmi les perspectives envisagées pour de futurs travaux, l'intégration de nouvelles variables telles que l'état de la route pourrait permettre d'affiner considérablement les prédictions. De plus, l'exploration de modèles hybrides combinant différentes approches pourrait représenter une voie prometteuse pour renforcer la généralisation et la robustesse des résultats.

En conclusion, cette étude contribue à une meilleure compréhension et modélisation du temps de conduite restant dans les véhicules électriques. Les résultats obtenus confirment la pertinence des approches d'apprentissage automatique pour la gestion énergétique intelligente et ouvrent la voie à des solutions plus précises, flexibles et fiables, essentielles au développement d'une mobilité électrique durable et optimisée.

LISTE DE RÉFÉRENCES

- [1] D. Chatroux, "Accumulateurs Lithium-ion et véhicules électriques," présenté à SGE 2020 - Symposium de Génie Electrique, Nantes, France, 2021-07-06 2021. [En ligne]. Disponible: <https://cea.hal.science/cea-03344278>
- [2] M. Muratori, M. Alexander, D. Arent, M. Bazilian, P. Cazzola, E. M. Dede, J. Farrell, C. Gearhart, D. Greene, A. Jenn, M. Keyser, T. Lipman, S. Narumanchi, A. Pesaran, R. Sioshansi, E. Suomalainen, G. Tal, K. Walkowicz, et J. Ward, "The rise of electric vehicles—2020 status and future expectations," *Progress in Energy*, vol. 3, no. 2, p. 022002, 2021/03/25 2021. doi: 10.1088/2516-1083/abe0ad.
- [3] J. Van Mierlo, M. Bercibar, M. El Baghdadi, C. De Cauwer, M. Messagie, T. Coosemans, V. A. Jacobs, et O. Hegazy, "Beyond the State of the Art of Electric Vehicles: A Fact-Based Paper of the Current and Prospective Electric Vehicle Technologies," *World Electric Vehicle Journal*, vol. 12, no. 1 2021. doi: 10.3390/wevj12010020.
- [4] H. El Jebbari, "Contribution à l'estimation de l'état de santé en vue de la prédiction de la durée de vie utile résiduelle des batteries au lithium ion. Application : véhicules électriques," 2021. [En ligne]. Disponible: <http://www.theses.fr/2021NORMC227/document>
- [5] T. Rahman et T. Alharbi, "Exploring Lithium-Ion Battery Degradation: A Concise Review of Critical Factors, Impacts, Data-Driven Degradation Estimation Techniques, and Sustainable Directions for Energy Storage Systems," *Batteries*, vol. 10, no. 7 2024. doi: 10.3390/batteries10070220.
- [6] A. T. Yapıcı, N. Abut, et T. Erfidan, "Comparing the Effectiveness of Deep Learning Approaches for Charging Time Prediction in Electric Vehicles: Kocaeli Example," *Energies*, vol. 18, no. 8 2025. doi: <https://doi.org/10.3390/en18081961>.
- [7] T. Montes, F. Pinsach Batet, L. Igualada, et J. Eichman, "Degradation-conscious charge management: Comparison of different techniques to include battery degradation in Electric Vehicle Charging Optimization," *Journal of Energy Storage*, vol. 88, p. 111560, 2024/05/30/ 2024. doi: <https://doi.org/10.1016/j.est.2024.111560>.
- [8] L. Zhao, W. Yao, Y. Wang, et J. Hu, "Machine Learning-Based Method for Remaining Range Prediction of Electric Vehicles," *IEEE Access*, vol. 8, pp. 212423-212441, 2020. doi: 10.1109/ACCESS.2020.3039815.
- [9] N. Wang, Y. Lyu, Y. Zhou, J. Luan, Y. Li, et C. Zheng, "A hybrid framework for remaining driving range prediction of electric taxis," *Sustainable Energy Technologies and Assessments*, vol. 67, p. 103832, 2024/07/01/ 2024. doi: 10.1016/j.seta.2024.103832.
- [10] W. Achariyaviriya, W. Wongsapai, K. Janpoom, T. Katongtung, Y. Mona, N. Tippayawong, et P. Suttakul, "Estimating Energy Consumption of Battery Electric Vehicles Using Vehicle Sensor Data and Machine Learning Approaches," *Energies*, vol. 16, no. 17 2023. doi: <https://doi.org/10.3390/en16176351>.
- [11] M. A. Eissa et P. Chen, "An Efficient Hybrid Deep Learning Approach for Accurate Remaining EV Range Prediction," Dans *2023 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, 2023, pp. 430-435.
- [12] S. Huang, Z. He, et X. Li, "A Method of SOC Estimation for Electric Vehicle Based on Limited Information," Dans *2020 IEEE International Conference on Networking, Sensing and Control (ICNSC)*, 2020, pp. 1-4.

- [13] H. Min, Y. Yan, W. Sun, Y. Yu, R. Jiang, et F. Meng, "Construction and Estimation of Battery State of Health Using a De-LSTM Model Based on Real Driving Data," *Energies*, vol. 16, no. 24 2023. doi: 10.3390/en16248088.
- [14] M. Ghalkhani et S. Habibi, "Review of the Li-Ion Battery, Thermal Management, and AI-Based Battery Management System for EV Application," *Energies*, vol. 16, no. 1 2023. doi: 10.3390/en16010185.
- [15] J. Liu, S. Yadav, M. Salman, S. Chavan, et S. C. Kim, "Review of thermal coupled battery models and parameter identification for lithium-ion battery heat generation in EV battery thermal management system," *International Journal of Heat and Mass Transfer*, vol. 218, p. 124748, 2024/01/01/ 2024. doi: <https://doi.org/10.1016/j.ijheatmasstransfer.2023.124748>.
- [16] K. Laadjal et A. J. M. Cardoso, "Estimation of Lithium-Ion Batteries State-Condition in Electric Vehicle Applications: Issues and State of the Art," *Electronics*, vol. 10, no. 13 2021. doi: 10.3390/electronics10131588.
- [17] J. Hong, Z. Wang, W. Chen, L. Wang, P. Lin, et C. Qu, "Online accurate state of health estimation for battery systems on real-world electric vehicles with variable driving conditions considered," *Journal of Cleaner Production*, vol. 294, p. 125814, 2021/04/20/ 2021. doi: <https://doi.org/10.1016/j.jclepro.2021.125814>.
- [18] D. V. Pelegov et J.-J. Chanaron, "Electric Car Market Analysis Using Open Data: Sales, Volatility Assessment, and Forecasting," *Sustainability*, vol. 15, no. 1 2023. doi: <https://doi.org/10.3390/su15010399>.
- [19] N. O. Kapustin et D. A. Grushevenko, "Long-term electric vehicles outlook and their potential impact on electric grid," *Energy Policy*, vol. 137, p. 111103, 2020/02/01/ 2020. doi: 10.1016/j.enpol.2019.111103.
- [20] F. Li, H. Feng, Y. Min, Y. Zhang, H. Zuo, F. Bai, et Y. Zhang, "Prediction of lithium-ion battery degradation trajectory in electric vehicles under real-world scenarios," *Energy*, vol. 317, p. 134663, 2025/02/15/ 2025. doi: <https://doi.org/10.1016/j.energy.2025.134663>.
- [21] A. Masakure, A. Gill, et M. Singh, "The Impact of Battery Charging and Discharging Current Limits on EV Battery Degradation and Safety," Dans *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)*, 2023, pp. 1-5.
- [22] N. Xu, X. Li, F. Yue, Y. Jia, Q. Liu, et D. Zhao, "An Eco-Driving Evaluation Method for Battery Electric Bus Drivers Using Low-Frequency Big Data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 9, pp. 9296-9308, 2023. doi: 10.1109/TITS.2023.3267187.
- [23] X. Ji, Y. Dang, M. Song, A. Liu, H. Zhao, et T. Jiang, "A universal method for seizure onset zone localization in focal epilepsy using standard deviation of spike amplitude," *Epilepsy Research*, vol. 208, p. 107475, 2024/12/01/ 2024. doi: <https://doi.org/10.1016/j.eplepsyres.2024.107475>.
- [24] S. Zhao, C. Ma, et Z. Cao, "Improved Multi-Objective Strategy Diversity Chaotic Particle Swarm Optimization of Ordered Charging Strategy for Electric Vehicles Considering User Behavior," *Energies*, vol. 18, no. 3 2025. doi: <https://doi.org/10.3390/en18030690>.
- [25] T. Zhang, X. Zhang, O. Rubasinghe, Y. Liu, Y. H. Chow, H. H. C. Iu, et T. Fernando, "Long-Term Energy and Peak Power Demand Forecasting Based on Sequential-XGBoost," *IEEE*

- Transactions on Power Systems*, vol. 39, no. 2, pp. 3088-3104, 2024. doi: 10.1109/TPWRS.2023.3289400.
- [26] H. Hu, J. Liu, X. Zhang, et M. Fang, "An Effective and Adaptable K-means Algorithm for Big Data Cluster Analysis," *Pattern Recognition*, vol. 139, p. 109404, 2023/07/01/ 2023. doi: <https://doi.org/10.1016/j.patcog.2023.109404>.
- [27] G. L. V. Prasad, B. R. Teja, T. Haribabu, G. N. Pavani, D. Karunamma, et K. Vivek, "A Hybrid Time Series Rainfall Prediction Model Using Neural Prophet and LS TM," Dans *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, 2023, pp. 1582-1587.
- [28] F. Liu et Y. Deng, "Determine the Number of Unknown Targets in Open World Based on Elbow Method," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 5, pp. 986-995, 2021. doi: 10.1109/TFUZZ.2020.2966182.
- [29] R. K. Gupta, M. R. Sonawane, T. Krishnan, K. K. P. Kumar, et S. Kumar, "Predictive Modelling of Photocatalytic Wastewater Treatment Using Random Forest and XGBoost on Mn-Doped BiFeO₃," Dans *2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, 2025, vol. 3, pp. 1-6.
- [30] M. Gupta, J. Mittal, et A. Tomar, "Predictive Performance of EV Charging Behaviour in COVID-19," Dans *2023 IEEE 3rd International Conference on Sustainable Energy and Future Electric Transportation (SEFET)*, 2023, pp. 1-6.
- [31] R. Sun, J. Chen, et C. Piao, "Battery health features extraction and state of health estimation based on real-time online vehicle driving data," *Journal of Power Sources*, vol. 645, p. 236784, 2025/07/30/ 2025. doi: <https://doi.org/10.1016/j.jpowsour.2025.236784>.
- [32] H. Liu, Q. Xiao, Z. Jiao, J. Meng, Y. Mu, K. Hou, X. Yu, S. Guo, et H. Jia, "LightGBM-Based Prediction of Remaining Useful Life for Electric Vehicle Battery under Driving Conditions," Dans *2020 IEEE Sustainable Power and Energy Conference (iSPEC)*, 2020, pp. 2577-2582.
- [33] A. A. Polat, S. Bozkurt Keser, İ. Sarıçiçek, et A. Yazıcı, "Analysis of Factors Affecting Electric Vehicle Range Estimation: A Case Study of the Eskisehir Osmangazi University Campus," *Sustainability*, vol. 17, no. 8 2025. doi: <https://doi.org/10.3390/su17083488>.
- [34] S. Koohfar, W. Woldemariam, et A. Kumar, "Performance Comparison of Deep Learning Approaches in Predicting EV Charging Demand," *Sustainability*, vol. 15, no. 5 2023. doi: <https://doi.org/10.3390/su15054258>.
- [35] W. Khan, "Short-Term Cost Forecasting of EV Battery Using LSTM Network," Dans *2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications(AEECA)*, 2020, pp. 372-376.
- [36] J. Baranowski, W. Bauer, N. Kashpruk, et M. Zagorowska, "Predicting system degradation using Bayesian time series models," Dans *2021 25th International Conference on Methods and Models in Automation and Robotics (MMAR)*, 2021, pp. 321-324.
- [37] A. L. Serrano, P. H. dos Santos Martins, G. F. Vergara, G. D. Bispo, G. A. Rodrigues, L. R. Mosquera, M. N. Oliveira, C. Neumann, M. G. Peixoto, et V. P. Gonçalves, "Forecasting Ethanol and Gasoline Consumption in Brazil: Advanced Temporal Models for Sustainable Energy Management," *Energies*, vol. 18, no. 6 2025. doi: <https://doi.org/10.3390/en18061501>.

- [38] S. Oyucu, F. Doğan, A. Aksöz, et E. Biçer, "Comparative Analysis of Commonly Used Machine Learning Approaches for Li-Ion Battery Performance Prediction and Management in Electric Vehicles," *Applied Sciences*, vol. 14, no. 6 2024. doi: <https://doi.org/10.3390/app14062306>.
- [39] D. Adhya, A. Pal, A. K. Chakraborty, et A. Bhattacharya, "Machine Learning Application for Prediction of EV Charging Demand for the Scenario of Agartala, India," Dans *2022 4th International Conference on Energy, Power and Environment (ICEPE)*, 2022, pp. 1-5.
- [40] Í. Viana dos Santos Santana, A. Cm da Silveira, Á. Sobrinho, L. Chaves e Silva, L. Dias da Silva, D. F. S. Santos, E. C. Gurjão, et A. Perkusich, "Classification Models for COVID-19 Test Prioritization in Brazil: Machine Learning Approach," *J Med Internet Res*, vol. 23, no. 4, p. e27293, 2021/4/8 2021. doi: <https://doi.org/10.2196/27293>.
- [41] Y. Gebreyesus, D. Dalton, S. Nixon, D. De Chiara, et M. Chinnici, "Machine Learning for Data Center Optimizations: Feature Selection Using Shapley Additive exPlanation (SHAP)," *Future Internet*, vol. 15, no. 3 2023. doi: <https://doi.org/10.3390/fi15030088>.

CERTIFICATION ÉTHIQUE

Ce projet ne nécessitait pas d'approbation éthique.