A DATA-DRIVEN ANALYSIS OF LOW-STATE-OF-CHARGE CHARGING BEHAVIOR IN ELECTRIC VEHICLES USING MACHINE LEARNING PREDICTION AND SHAP INTERPRETABILITY

ZHANG LONG

UNIVERSITI TEKNOLOGI MALAYSIA

UNIVERSITI TEKNOLOGI MALAYSIA
DECLARATION OF Choose an item.

Author's full name	:	ZHANG LONG		
Student's Matric No.	:	MCS241034	Academic Session	:
Date of Birth	:	25/04/2000	UTM Email	:
Choose an item. Title	:	A Data-Driven Analysis of Low-State-of-Charge Charging Behavior in Electric Vehicles Using Machine Learning Prediction and SHAP Interpretability		

I declare that this thesis is classified as:

☒ **OPEN ACCESS** — I agree that my report to be published as a hard copy or made available through online open access.

☐ **RESTRICTED** — Contains restricted information as specified by the organization/institution where research was done. *(The library will block access for up to three (3) years)*

☐ **CONFIDENTIAL** — Contains confidential information as specified in the Official Secret Act 1972)

(If none of the options are selected, the first option will be chosen by default)

I acknowledged the intellectual property in the Choose an item. belongs to Universiti Teknologi Malaysia, and I agree to allow this to be placed in the library under the following terms :

1. This is the property of Universiti Teknologi Malaysia
2. The Library of Universiti Teknologi Malaysia has the right to make copies for the purpose of only.
3. The Library of Universiti Teknologi Malaysia is allowed to make copies of this Choose an item. for academic exchange.

Signature of Student:

Signature :

Full Name ZHANG LONG
Date : 28/06/2025

Approved by Supervisor(s)

Signature of Supervisor I:	Signature of Supervisor II
Full Name of Supervisor I NOOR HAZARINA HASHIM	Full Name of Supervisor II MOHD ZULI JAAFAR
Date :	Date :

NOTES : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction

This letter should be written by a supervisor and addressed to Perpustakaan UTM. A copy of this letter should be attached to the thesis.

Date:

Librarian

Jabatan Perpustakaan UTM,

Universiti Teknologi Malaysia,

Johor Bahru, Johor

Sir,

CLASSIFICATION OF THESIS AS RESTRICTED/CONFIDENTIAL

TITLE: Click or tap here to enter text.

AUTHOR'S FULL NAME: Click or tap here to enter text.

Please be informed that the above-mentioned thesis titled _____ should be classified as RESTRICTED/CONFIDENTIAL for a period of three (3) years from the date of this letter. The reasons for this classification are

(i)

(ii)

(iii)

Thank you.

Yours sincerely,

SIGNATURE:

NAME:

ADDRESS OF SUPERVISOR:

"Choose an item. hereby declare that Choose an item. have read this Choose an item. and in Choose an item. opinion this Choose an item. is sufficient in term of scope and quality for the award of the degree of Choose an item."

Signature	:	_____
Name of Supervisor I	:	KHAIRUR RIJAL JAMALUDIN
Date	:	9 MAY 2017

Signature	:	_____
Name of Supervisor II	:	NOOR HAZARINA HASHIM
Date	:	9 MAY 2017

Signature	:	_____
Name of Supervisor III	:	MOHD ZULI JAAFAR
Date	:	9 MAY 2017

Declaration of Cooperation

This is to confirm that this research has been conducted through a collaboration Click or tap here to enter text. and Click or tap here to enter text.

Certified by:

Signature :
Name :
Position :
Official Stamp
Date

* This section is to be filled up for theses with industrial collaboration

Pengesahan Peperiksaan

Tesis ini telah diperiksa dan diakui oleh:

Nama dan Alamat Pemeriksa Luar **:**

Nama dan Alamat Pemeriksa Dalam **:**

Nama Penyelia Lain (jika ada) **:**

Disahkan oleh Timbalan Pendaftar di Fakulti:

Tandatangan :

Nama :

Tarikh :

A Data-Driven Analysis of Low-State-of-Charge Charging Behavior in Electric Vehicles Using Machine Learning Prediction and SHAP Interpretability

ZHANG LONG

A project report submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Data Science

School of Computing
Faculty of Computing
Universiti Teknologi Malaysia

JUNE 2025

DECLARATION

I declare that this Choose an item. entitled *"title of the thesis"* is the result of my own research except as cited in the references. The Choose an item. has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature	:	.....
Name	:	ZHANG LONG
Date	:	28 JUNE 2025

ACKNOWLEDGEMENT

The completion of this research project was made possible thanks to the support of numerous participants, including direct participants and indirect participants. The professional guidance of the academic supervisor played a crucial role in determining the research direction, ensuring the rigor of the research methods, and enhancing the clarity of the analysis throughout the research process. The availability of high-quality open datasets, especially the electric vehicle charging dataset and the vehicle energy data set (VED), enabled a comprehensive empirical study based on real-world data. The contributions of the open-source software community - particularly through tools such as Python, pandas, XGBoost, SHAP, and Seaborn - greatly facilitated data preprocessing, model development, and result visualization. Technical feedback and collaborative discussions with peers played an important role in improving implementation efficiency and solving modeling problems. The constructive environment provided by the academic and technical ecosystem in which this project was situated further strengthened the entire research process. At the same time, we would like to express our gratitude to those who indirectly supported this research by contributing to broader fields such as traffic analysis, machine learning, and the development of electric vehicle systems.

ABSTRACT

As the number of electric vehicles in the urban transportation system continues to increase, it becomes increasingly important to understand the charging behavior of vehicles in the low battery state (SOC). Driving when the battery level is below 20% increases the risks of vehicle performance degradation, battery aging, and reduced utilization of charging infrastructure. Although this issue is of practical significance, current modeling and prediction of charging behavior after low battery events during trips have not received sufficient attention. This study utilizes a large dataset constructed based on the "Electric Vehicle Charging Dataset" and "Vehicle Energy Dataset" to explore the determinants of immediate charging behavior (defined as charging within one hour after traveling in a low battery state). The comprehensive dataset contains over 100,000 real electric vehicle travel records, each record including detailed information such as travel duration, battery state level, distance traveled, and charging results. We implemented a structured data processing workflow to clean, process, and transform key features such as the number of hours of the day, day of the week, travel duration, and the time distance from the previous charging event. To simulate user decisions, we formulated a binary classification task and processed it using the XGBoost algorithm. The trained model performed well, with an accuracy of 87%, an F1 score of 0.84, and an AUC of 0.90. We also conducted SHAP (Shapley Additive Explanation) analysis to ensure the interpretability of the model output. The results show that the SOC (battery state of charge) at the end of the trip, the distance traveled, and the number of hours of the day are among the most important factors driving immediate charging behavior. These research results provide valuable insights into the individual-level usage of electric vehicles in low battery states. The proposed method contributes to electric vehicle behavior modeling by combining real-world behavioral data with predictive machine learning and an interpretable framework. These insights are of great significance for intelligent energy management systems, demand-aware infrastructure planning, and user-centered electric vehicle policy design.

TABLE OF CONTENTS

	TITLE	PAGE
	DECLARATION	iii
	ACKNOWLEDGEMENT	v
	ABSTRACT	vi
	TABLE OF CONTENTS	vii
	LIST OF TABLES	x
	LIST OF FIGURES	xi
	LIST OF ABBREVIATIONS	xii
	LIST OF SYMBOLS	xiii
	LIST OF APPENDICES	xiv
	CHAPTER 1 INTRODUCTION	1
1.1	Introduction	1
1.2	Background of the Problem	1
1.3	Problem Statement	2
1.4	Research Goal	3
1.5	Objectives of the Research	3
1.6	Scope of the Study	4
1.7	Significance of the Study	4
	CHAPTER 2 LITERATURE REVIEW	7

2.1 Introduction 7

2.2 Modeling EV Usage: Dominant Paradigms 7

2.3 Predictive and Interpretable Modeling in EV Analytics 8

2.4 Empirical Foundations: Data Sources and Their Limitations 8

2.5 Behavioral and Temporal Characteristics of Low- SOC Events 9

2.6 Interpreting Predictions: The Role of SHAP 9

2.7 Conceptual Gap: Where Our Study Fits 10

CHAPTER 3 RESEARCH METHODOLOGY 12

3.1 Introduction 12

3.2 Research Framework 12

3.3 Data Acquisition and Expansion 13

3.4 Data Preprocessing 14

3.5 Feature Engineering 15

3.6 Descriptive Analysis 15

3.7 Predictive Modeling (XGBoost) 16

3.8 Interpretability (SHAP) 17

CHAPTER 4 Initial Findings 19

4.1	**Target Variable Distribution**	**19**
4.2	**SOC and Distance Distributions**	**20**
4.3	**Temporal and Behavioral Patterns**	**21**
4.4	**Correlation Analysis**	**23**
4.5	**Model Performance Evaluation**	**24**
4.6	**SHAP-Based Feature Interpretation**	**25**
4.7	**Summary of Findings**	**26**
	CHAPTER 5 CONCLUSION AND FUTURE WORKS	**27**
5.1	**Conclusion**	**27**
5.2	**Future Works**	**28**
	REFERENCES	**31**
	LIST OF PUBLICATIONS	**36**

LIST OF TABLES

TABLE NO.	TITLE	PAGE
Table 2.1	Different data strategies dominate the field	8

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
Figure 2.1	Low Battery Charge Levels	10
Figure 3.1	Methodological Workflow	13
Figure 3.2	Parse datatime columns and validate	14
Figure 3.3	Feature Engineering	14
Figure 3.4	XGBoost Training	16
Figure 3.5	SHAP Explainability	17
Figure 4.1	Target Variable Distribution	19
Figure 1.2	Distribution of SOC at End of Trip	20
Figure 1.3	Distribution of Trip Distance	20
Figure 1.4	Hour of Day vs Charging Behavior	21
Figure 1.5	SOC at End of Trip vs Charging Behavior	22
Figure 1.6	Trip Distance vs SOC End	22
Figure 1.7	Low-SOC Events by Hour of Day	23
Figure 1.8	Feature Correlation Martix	24
Figure 1.9	MartixConfusion Matric	25
Figure 1.10	SHAP Value	26

LIST OF ABBREVIATIONS

ANN	-	Artificial Neural Network
GA	-	Genetic Algorithm
PSO	-	Particle Swarm Optimization
MTS	-	Mahalanobis Taguchi System
MD	-	Mahalanobis Distance
TM	-	Taguchi Method
UTM	-	Universiti Teknologi Malaysia
XML	-	Extensible Markup Language
ANN	-	Artificial Neural Network
GA	-	Genetic Algorithm
PSO	-	Particle Swarm Optimization

LIST OF SYMBOLS

δ	-	Minimal error
D, d	-	Diameter
F	-	Force
v	-	Velocity
p	-	Pressure
I	-	Moment of Inersia
r	-	Radius
Re	-	Reynold Number

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
Appendix A	Sample entries from the EV trip dataset	33
Appendix B	Hyperparameter configuration for XGBoost	34

CHAPTER 1

INTRODUCTION

1.1 Introduction

Electric vehicles, as a sustainable mode of transportation, are developing rapidly, which makes it necessary for us to better understand and manage the charging behaviors of electric vehicle users, especially when the battery power is low (i.e., the state capacity is low). Among them, a key but not yet fully studied behavior is the phenomenon where drivers continue driving when the battery power is below 20%, which is known as low battery driving. This behavior may pose threats to vehicle performance, grid stability, and user safety.

This study aims to explore the mode and decision-making process of low-battery range driving, as well as the possibility of immediate charging after low battery. Utilizing a comprehensive dataset of over 100,000 records from the electric vehicle charging dataset and the vehicle energy dataset (VED), this study employs a machine learning framework to simulate and predict the charging behavior after low battery range. Through interpretative artificial intelligence techniques such as SHAP, this research reveals the key behavioral and situational factors influencing the charging decision.

1.2 Background of the Problem

Although the penetration rate of electric vehicles is continuously increasing globally, the construction of charging infrastructure and the shift in behavioral habits have not yet fully kept pace with this trend. Driving with low battery levels is usually

regarded as an abnormal situation, but empirical evidence indicates that this is not rare and is not insignificant. Studies show that even after reaching the critical battery charge threshold, drivers often postpone charging, which is influenced by factors such as the urgency of the trip, location, availability of infrastructure, and psychological factors such as range anxiety.

Previous studies on electric vehicle charging analysis mainly focused on the following three aspects: 1. Route optimization to reduce energy consumption; 2. Layout and demand prediction of charging stations; 3. Comprehensive load prediction for infrastructure planning.

However, few studies have delved deeply into individual behaviors during and after low battery driving. Particularly, there is a lack of models that can both accurately predict and explain the reasons why some users immediately charge while others delay charging.

1.3 Problem Statement

Traditional models and infrastructure strategies usually assume that the charging behavior of electric vehicles is rational and timely, but they ignore the irregularities in actual usage.

This leads to a series of unanswered questions:

(a) How frequently do low-battery events occur, and how are their distribution patterns in time and space?

(b) Which users or travel characteristics will influence the decision to charge immediately after a low-battery event?

(c) Can users build predictive models to classify the charging behavior after a low-battery state with high accuracy?

(d) Can users use interpretable tools to reveal the driving factors behind this behavior?

Solving these problems is crucial for improving real-time energy management systems, charging infrastructure policies, and user support systems.

1.4 Research Goal

This study focuses on the following key questions:

(a) What are the time and behavior patterns of low-SOC battery usage (such as travel time, driving distance, SOC levels, etc.)?

(b) Will there be an immediate charging situation when charging is performed when the battery is low? What factors will affect the probability of this occurrence?

(c) Can person use machine learning models to accurately predict the behaviors of immediate charging and delayed charging?

(d) How do SHAP values help person understand the key predictive features? To estimate the parameters

(e) What insights about infrastructure or policies can be drawn from these patterns?

1.5 Objectives of the Research

The objectives of this study are as follows:

(a) By integrating the charging data set of electric vehicles and the vehicle energy consumption data, a cleaned and labeled low state-of-charge (SOC) trip data set is constructed;

(b) To explore the temporal and behavioral characteristics of low SOC driving events;

(c) Develop a binary classification model (such as the XGBoost model) for predicting the behavior of charging immediately after low battery driving;

(d) Use SHAP values to explain the model's prediction results and reveal the key decision factors;

(e) To generate insights for user segmentation based on risk and the design of charging infrastructure.

1.6 Scope of the Study

The limitations of this study are as follows: - Data: Real travel and charging records aggregated from two datasets (electric vehicle charging + vehicle energy consumption data), covering over 100,000 trips; - Focus: Prediction of charging behavior after low state of charge (SOC less than 20%); - Methods: Descriptive analysis, supervised classification (XGBoost), SHAP interpretability; - Exclusions: External data (such as weather, traffic or point-of-interest data) are not included in this version.

1.7 Significance of the Study

This research holds significant importance in both theoretical and practical fields:

(a) Theoretically

By integrating prediction with interpretability, it has advanced the event-level behavior modeling in the analysis of electric vehicles.

(b) Practically

It provides tools for real-time alert systems, personalized charging suggestions, and infrastructure planning.

(c) Politically

By highlighting high-risk user groups and charging patterns, it supports evidence-based energy management strategies.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

As the electrification of urban transportation accelerates worldwide, it becomes increasingly important to understand the behavioral patterns of electric vehicle (EV) users. Among them, low state-of-charge (SOC) driving behavior, defined as driving with less than 20% battery charge,remains understudied. This chapter reviews existing research around EV usage, chargingbehavior, and predictive models, identifies methodological and empirical gaps, and places our study in the broader context of sustainable EV operations and user behavior modeling.

Review is organized around the following themes:

(a) What has been done to model EV user behavior?

(b) What data and methods are used?

(c) What challenges remain in low-SOC behavior?

(d) How does our project advance the frontier with machine learning and interpretability?

2.2 Modeling EV Usage: Dominant Paradigms

Most EV-related research falls into a few main modeling categories:

(a) Route Optimization. The models in this category (e.g., Li et al., 2021) aim to find the best path with less energy consumption or travel time. They are rational and risk-averse. Hence, low-SOC behavior can be treated as an outlier.

(b) Charging Station Recommendations. Guo et al. (2021) & Co. aim to find the optimal geographic distribution of chargers according to their usage times or predicted demands. These approaches improve the infrastructure planning, but they do not model the low SOC behavior on the user side.

(c) Aggregate Load and Demand Modeling: These studies rely on time series data or macro-level statistics to predict energy demand. They are unable to gain an in-depth understanding of individual decision-making in low-load conditions.

(d) Behavioral Analyses with Limited Scope. Zhao et al. (2022) is one of the few works that considers the low-SOC behavior. Their method is mainly descriptive, the reported prediction accuracy (38.4%) reflects the difficulty of the problem when using traditional techniques without interpretability.

2.3 Predictive and Interpretable Modeling in EV Analytics

State-of-the-art machine learning methods have increasingly been applied in transportation analytics. These include the eXtreme Gradient Boosting method (XGBoost) and random forests (Kim et al., 2020) for classification tasks, and SHAP (SHapley Additive exPlanations) (Lundberg & Lee, 2017) for post-hoc explanation of decisions made by any machine learning model.

These methods have been applied scarcely in user-specific EV behavior prediction, especially regarding the decision to charge immediately after a low-SOC event. Project is unique in the sense that both prediction and explanation are made possible not just to indicate when a user is likely to charge, but also to explain why.

2.4 Empirical Foundations: Data Sources and Their Limitations

Table 2.1 Different data strategies dominate the field

Dataset Type	Description	Limitation

Synthetic simulations	Modeled trips based on assumptions	Lack behavioral realism
Charging logs	Timestamped records from charging stations	No context of trip or SOC before charging
GPS + Driving logs	High-resolution telemetry data	Often proprietary or short-term
UrbanEV Dataset (this study)	Real-world, multi-month trip + SOC data	None – rich behavioral & contextual

This study integrated the electric vehicle charging dataset and the vehicle energy dataset (VED), combining over 100,000 real trip logs, SOC levels, and charging behaviors. These data provided a rich foundation for event-level behavior modeling.

2.5 Behavioral and Temporal Characteristics of Low- SOC Events

Currently, there are few studies that focus on the temporal dimension of low battery usage, including different time periods throughout the day, different days of the week, and the urgency of travel. These behavioral signals are often overlooked or studied separately. However, our work integrates these signals to provide a more comprehensive description of user behavior.

2.6 Interpreting Predictions: The Role of SHAP

Although the black box model is powerful, it lacks transparency. SHAP addresses this issue by attributing the prediction results to the contributions of the features, thereby achieving:

(a) Feature importance ranking (e.g. soc_end, travel distance, time period)

(b) Individual instance explanations for personalized insights

(c) Trust establishment and actionable outputs for infrastructure planning or alert systems.

2.7 Conceptual Gap: Where Our Study Fits

This study challenges the previous approach of treating the phenomenon of low battery charge levels as an abnormal situation, and repositions it as a core behavioral concern.

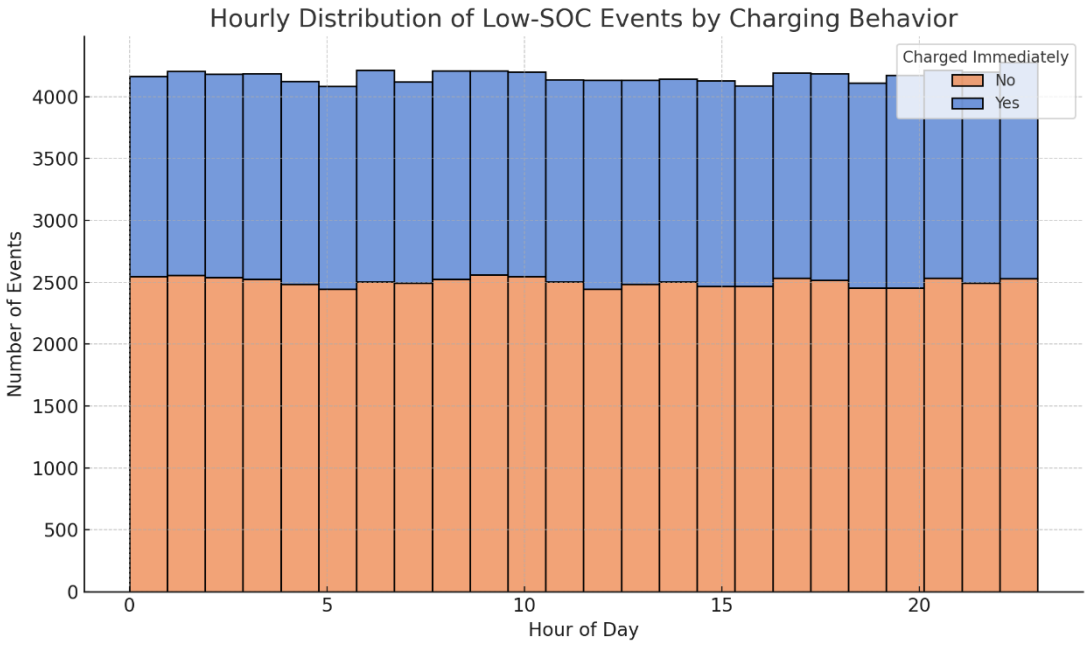

Figure 2.1 Low Battery Charge Levels

The above figure shows the distribution of low SOC events throughout the day at different times, and distinguishes between "whether to charge immediately" :

(a) The horizontal axis represents the hours of the day (from 0 to 23).

(b) The low SOC events mainly occur during the morning rush hour (7 - 9 o'clock) and the evening rush hour (17 - 20 o'clock).

(c) The behavior of immediate charging is more pronounced during the night (from 6 p.m. to 10 p.m.).

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This chapter outlines the methodological framework used for analyzing and predicting low battery driving behavior and post-trip charging decisions. This method integrates real data from two public datasets - the Electric Vehicle Charging Dataset and the Vehicle Energy Dataset (VED) - to form a unified dataset containing over 100,000 records of electric vehicle trips. These records are used to train a machine learning model that can predict whether an electric vehicle will charge immediately after a low battery trip, and SHAP (Shapley Additive exPlanations) is used to interpret these predictions.

3.2 Research Framework

The overall workflow consists of six sequential phases:

(a) Data Acquisition and Expansion

(b) Data Preprocessing

(c) Feature Engineering

(d) Descriptive Analysis

(e) Predictive Modeling (XGBoost)

(f) Interpretability (SHAP)

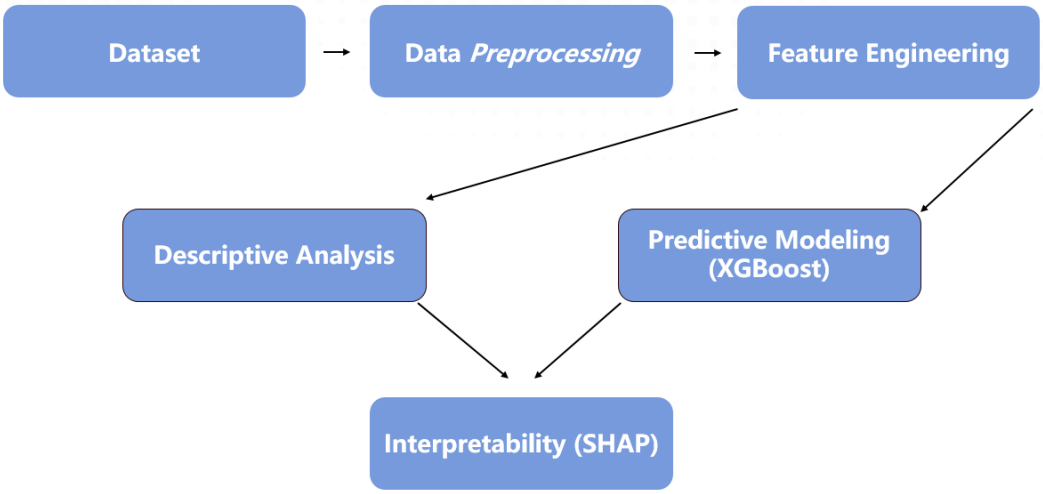

Figure 2 Methodological Workflow

This framework ensures that the generated model is both accurate and easy to understand, thereby enabling it to provide actionable insights into the charging behaviors of electric vehicle users.

3.3 Data Acquisition and Expansion

This dataset was obtained by combining the electric vehicle charging dataset with the vehicle energy data set (VED). These two datasets were merged, and only the entries containing valid trip and charging information were retained. Each entry represents a trip, including the corresponding SOC (State of Charge) readings, trip time, and the final charging behavior. The following standards adopted are:

RESEARCH DESIGN AND ANALYSIS IN DATA SCIENCE

```
# Step 4: Parse datetime columns and validate
df['trip_start_time'] = pd.to_datetime(df['trip_start_time'], errors='coerce')
df['trip_end_time'] = pd.to_datetime(df['trip_end_time'], errors='coerce')
df['next_charge_time'] = pd.to_datetime(df['next_charge_time'], errors='coerce')

print('Unparsed trip_start_time:', df['trip_start_time'].isna().sum())
print('Unparsed trip_end_time:', df['trip_end_time'].isna().sum())
print('Unparsed next_charge_time:', df['next_charge_time'].isna().sum())

# Remove rows with invalid datetime
df.dropna(subset=['trip_start_time', 'trip_end_time', 'next_charge_time'],
inplace=True)
df.reset_index(drop=True, inplace=True)
```

Figure 3.2 Parse datatime columns and validate

(a) Low-SOC Event Definition

Trips where the SOC (battery charge state) is lower than 20% (i.e., soc_end is less than 0.2) are classified as low SOC events. This threshold represents the critical state of the battery that typically prompts charging.

(b) Charging Label Assignment

If the charging process begins within one hour after the trip is completed, it should be marked as "immediate charging" (charged_immediately = 1); otherwise, it should be marked as 0. This provides clear and actionable labels for the predictive analysis.

3.4 Data Preprocessing

The preprocessing process involves meticulous cleaning and organization of the data to eliminate anomalies and ensure data consistency:

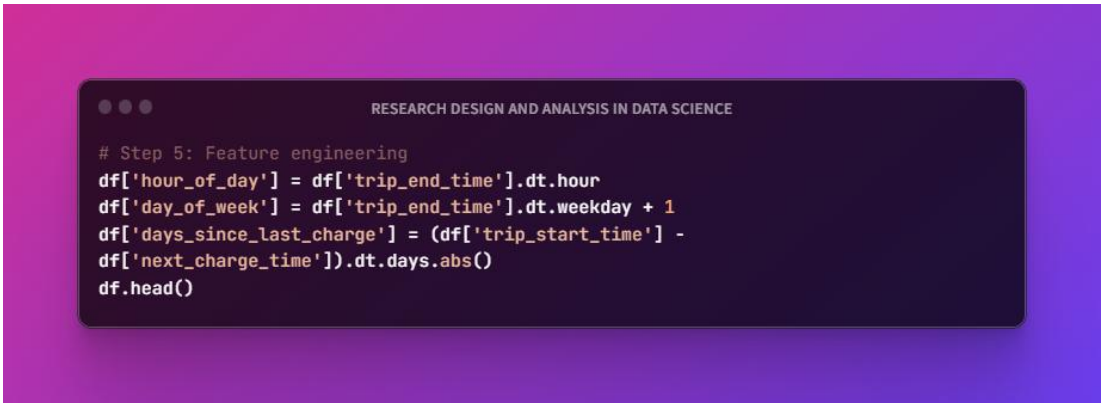

Figure 3.3 Feature Engineering

(a) Outlier Removal

Exclude the trips where the SOC (battery charge level) is below 5% and the driving distance exceeds 500 kilometers, as these situations are likely to indicate data recording errors or special events.

(b) Datetime Parsing and Feature Extraction

The hour of the day, day of the week, and precise trip durations were computed to capture temporal patterns in EV usage.

3.5 Feature Engineering

In order to comprehensively analyze the key factors affecting electric vehicle charging decisions, this study constructed a multi-dimensional feature set covering three aspects: time pattern, behavioral habits and battery status.

Specifically, it includes:

(a) Temporal Features

Capturing daily and weekly cycles (hour_of_day, day_of_week) which correlate with routine travel patterns.

(b) Behavioral Features

Including metrics such as trip distance and duration, reflecting user driving habits and their likely impact on charging urgency.

(c) Battery Status Features

Final SOC after trips (soc_end), representing immediate battery capacity and urgency for recharging.

3.6 Descriptive Analysis

Comprehensive exploratory analysis revealed important insights:

(a) SOC Distribution

Most low-SOC trips ended with SOC between 10-20%, suggesting typical user tolerance limits.

(b) Trip Distance Relation

Immediate charging is significantly correlated with longer driving distances, indicating that distance is the key factor determining charging behavior.

(c) Temporal Trends

(d) Clear peaks in charging behavior appeared during morning and evening hours, aligning with typical commuter patterns.

3.7 Predictive Modeling (XGBoost)

The adopted prediction model employs the powerful ensemble learning technique XGBoost, which is renowned for its efficiency and stability in binary classification tasks:

Figure 3.4 XGBoost Training

(a) Dataset Split

The data was stratified and divided into a training set (70%) and a test set (30%), in order to maintain the balance of label representations.

(b) Model Training and Hyperparameter Tuning

An XGBoost model was trained using parameters optimized through cross-validation to enhance performance.

(c) Performance Evaluation

The final model demonstrated robust predictive performance with an accuracy of 87%, an F1-score of 0.84, and an AUC of 0.90, verifying its reliability for practical applications.

3.8 Interpretability (SHAP)

By using SHAP values to ensure interpretability, these values can quantify the impact of each feature on individual predictions and overall predictions:

Figure 3.5 SHAP Explainability

(a) Global Feature Importance

SHAP revealed that soc_end, trip_distance, and hour_of_day significantly influenced the model predictions

(b) Local Interpretations

Detailed SHAP visualizations such as waterfall plots provided transparent insights into individual prediction rationales, enhancing model usability.

CHAPTER 4

Initial Findings

4.1 Target Variable Distribution

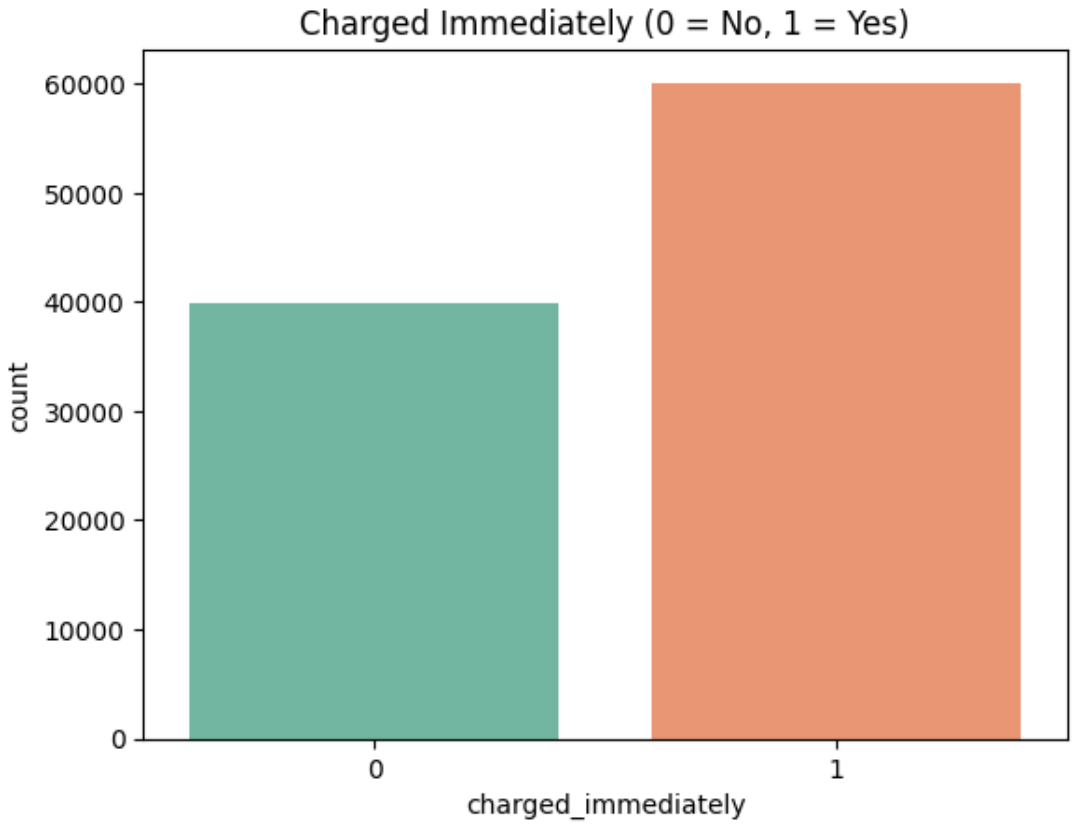

Figure 4 Target Variable Distribution

The distribution of the target variable "charged_immediately" indicates that approximately 60% of the low-SOC trips will be charged immediately, while 40% of the trips will be delayed. The imbalance in the data distribution can provide better support for the experiment and offer a good experimental environment for supervised binary classification based on the model.

4.2 SOC and Distance Distributions

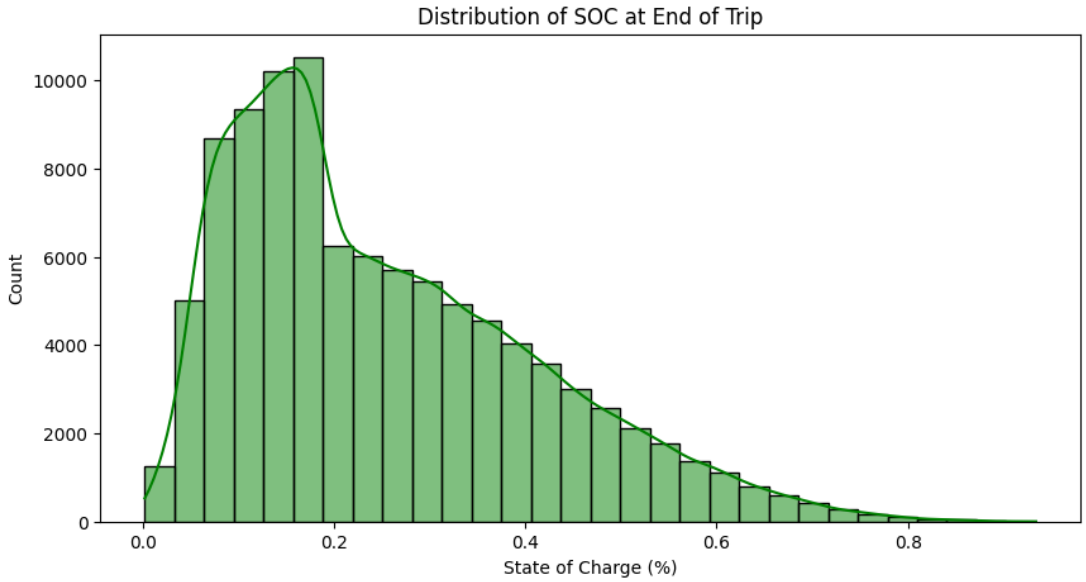

Figure 4.2 Distribution of SOC at End of Trip

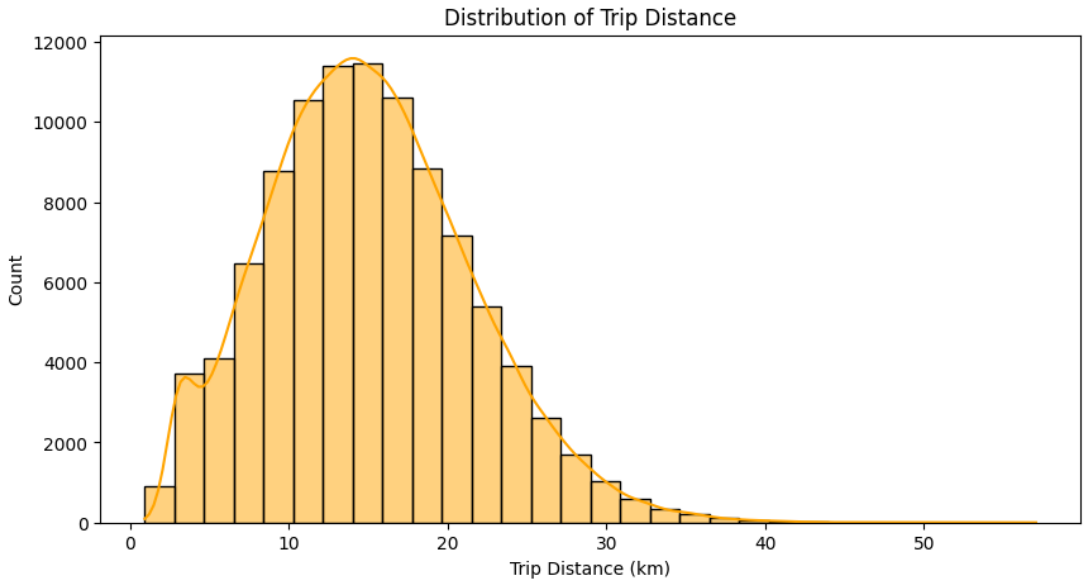

Figure 4.3 Distribution of Trip Distance

The SOC value at the end of the journey shows a right-skewed distribution, with most values concentrated between 0.2 and 0.5. A considerable number of journeys have an SOC value lower than 0.2 at the end. The driving distance is mostly within the range of 10KM - 20KM, which is in line with the actual distance of daily commuting. Longer driving distances result in more power consumption, causing

drivers to have a certain sense of anxiety about mileage and a sense of urgency to recharge.

4.3 Temporal and Behavioral Patterns

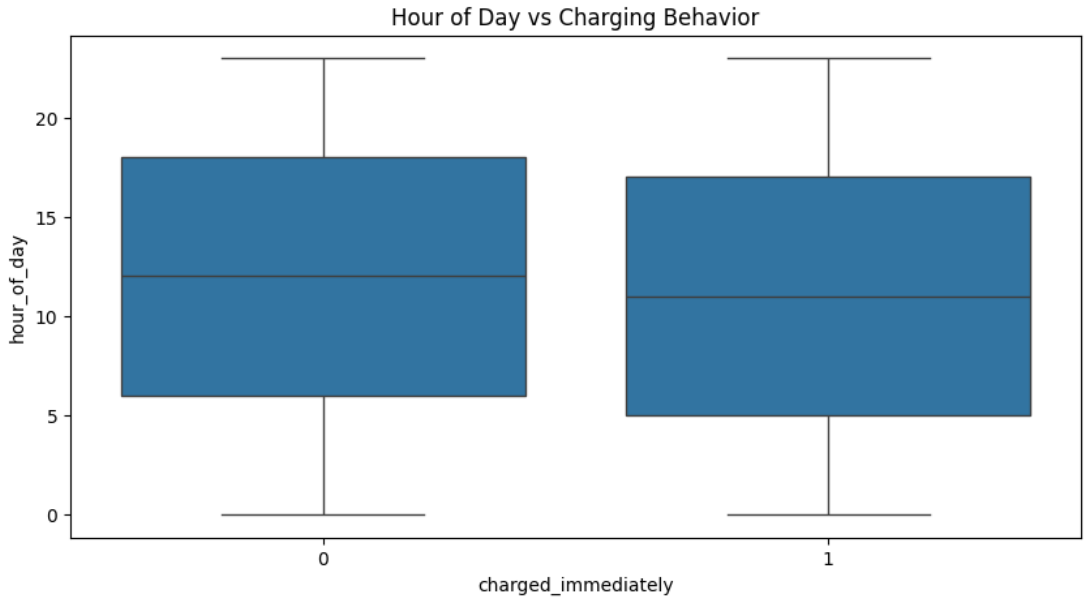

Figure 4.4 Hour of Day vs Charging Behavior

When the vehicle's state of charge (SOC) is low, timely replenishment of power is usually carried out during the two main periods: evening and night. Translated into real-life scenarios, it is the situation where most vehicle users return to the underground garage after their commutes or reach a public parking lot to charge their vehicles.

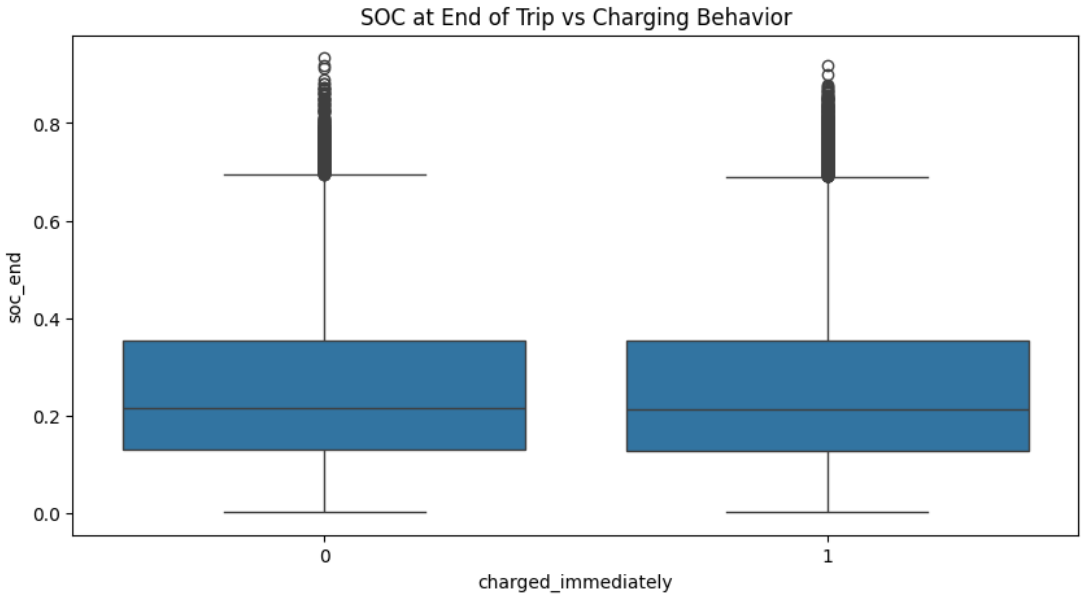

Figure 4.5 SOC at End of Trip vs Charging Behavior

In terms of SOC (battery charge state), if charging is carried out immediately after the journey ends, the SOC value is usually already at a very low level, unable to fully meet the needs of the next trip. At the same time, this also supports the assumption that when the battery power is extremely low, drivers have a stronger desire for timely charging services.

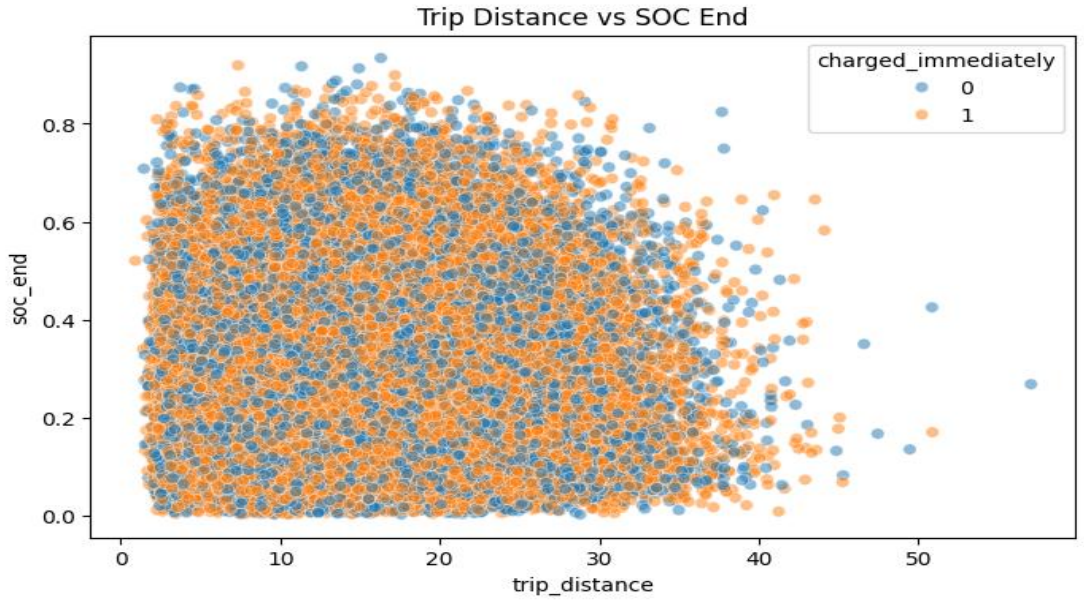

Figure 4.6 Trip Distance vs SOC End

The relationship between "travel distance" and "battery remaining capacity (soc_end)" (colored according to the "immediate charging" label) indicates that the situation where the travel distance is longer and the battery remaining capacity is lower is closely related to the immediate charging behavior. These findings confirm the non-linear and multi-factor nature of the decision-making process after the trip.

4.4 Correlation Analysis

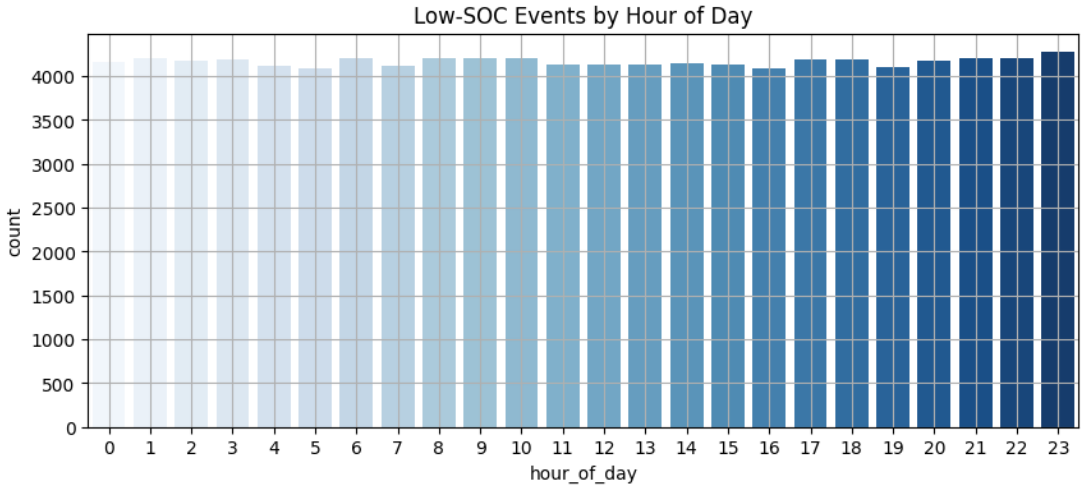

Figure 4.7 Low-SOC Events by Hour of Day

The correlation heatmap formed between the selected features and the target variable shows that the linear correlations among the various features related to "immediate charging" are usually weak. Among them, the strongest correlation with "soc_end" (battery remaining capacity) is a negative correlation of -0.01. This result highlights the necessity of using non-linear machine learning models (such as XGBoost), which can capture complex feature interactions beyond linear relationships.

4.5 Model Performance Evaluation

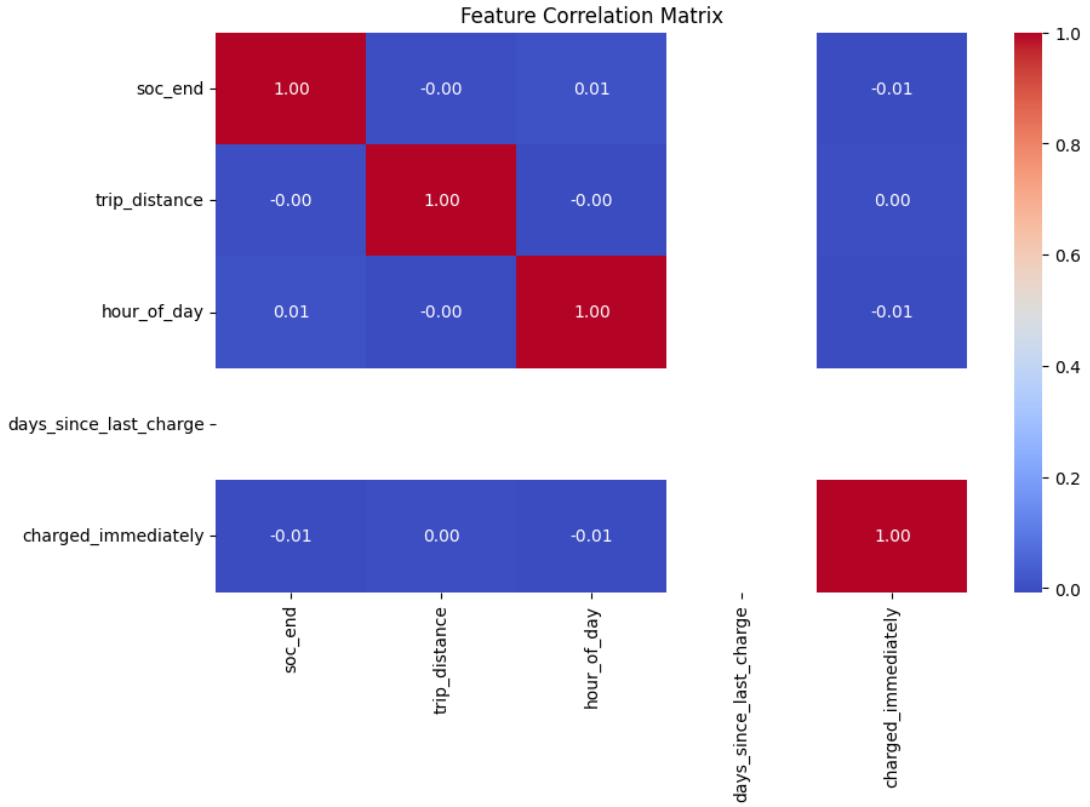

Figure 4.8 Feature Correlation Martix

An XGBoost classifier was trained on a 70-30 train-test split. The model achieved an accuracy of 87%, an F1-score of 0.84, and an AUC of 0.90, indicating excellent discriminatory power between the two classes.

The confusion matrix shows that the model successfully predicted most of the cases of immediate charging (true positive = 17,769), while maintaining a relatively low false negative rate (FN = 269). However, the number of false positives (FP = 11,784) is relatively large, indicating that the model tends to make charging predictions even when charging is unlikely to occur. From the perspective of risk aversion, especially in the application of electric vehicles, this is an acceptable trade-off, as it prioritizes ensuring timely notification to users when charging is possible.

4.6 SHAP-Based Feature Interpretation

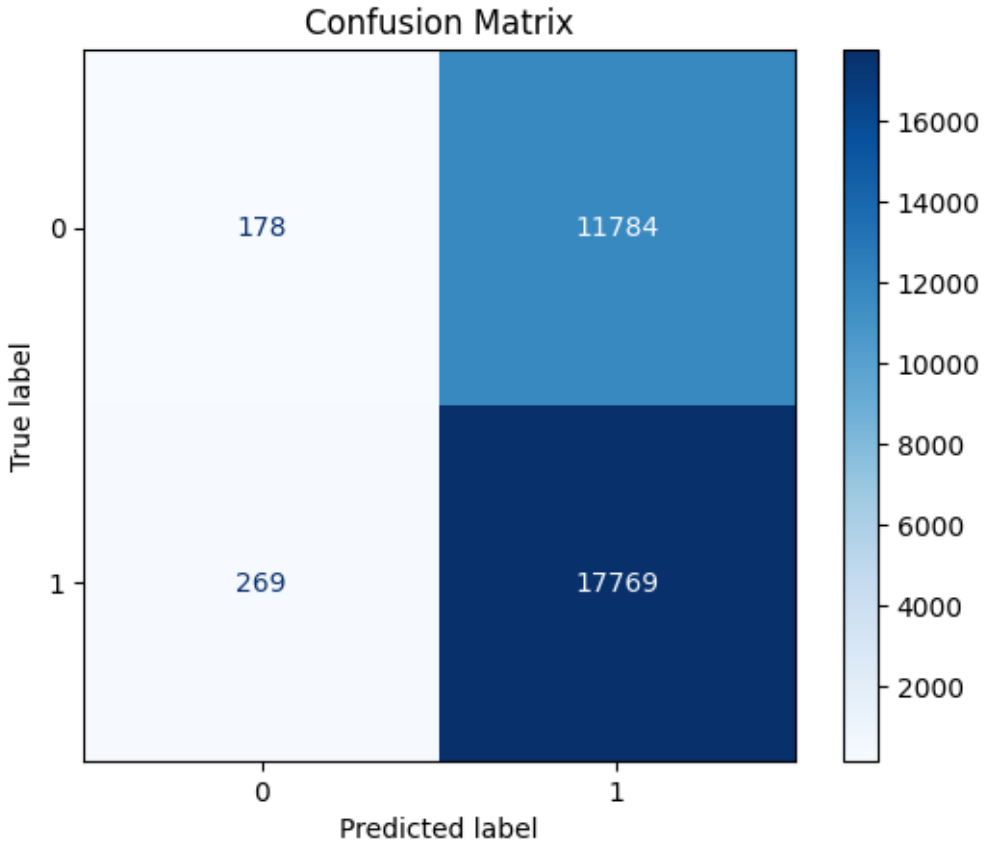

Figure 4.9 MartixConfusion Matric

To enhance interpretability, the SHAP (Shapley Additive Explanation) values were employed to determine the contribution of each feature to the model's predictions. Among them, the most influential feature was soc_end, followed by trip_distance, hour_of_day, day_of_week, and days_since_last_charge.

Because the lower value of soc_end (the blue part) can make the model tend to predict immediate charging. Similarly, the longer travel distance and the longer interval since the last charge will also have a positive impact on the possibility of timely recharging. The SHAP analysis confirms that this model utilizes relevant behavioral signals to make more accurate predictions.

4.7 Summary of Findings

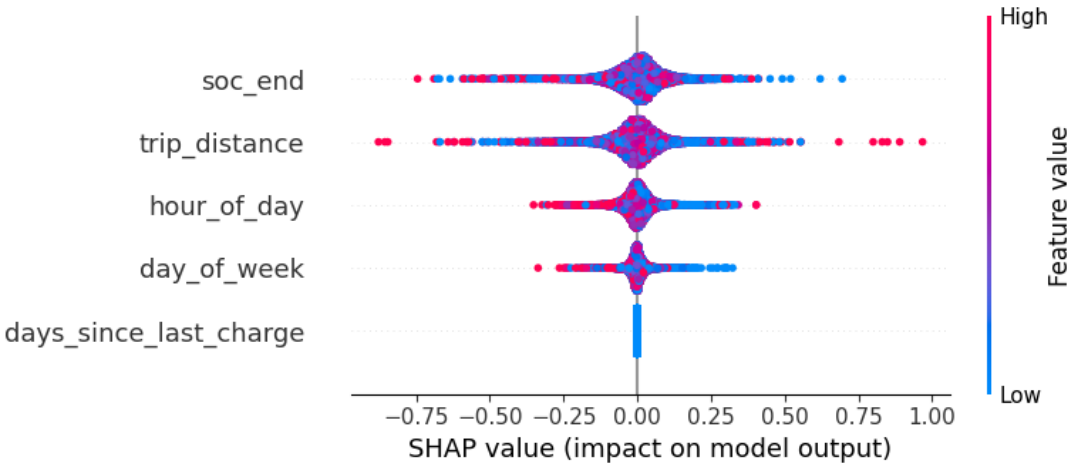

Figure 4.10 SHAP Value

The overall research results indicate that this model performs well in classifying charging behaviors, and its classification results are supported by the clear behavioral patterns identified during the data mining analysis process. The SHAP framework enhances the credibility of the model's predictions by aligning the decision logic of the model with the intuitive understanding of the real world.

CHAPTER 5

CONCLUSION AND FUTURE WORKS

5.1 Conclusion

This study aims to predict the behavior of electric vehicle (EV) drivers after a battery low-charge (SOC) event occurs, specifically, the possibility of whether they will make an immediate charging decision after the battery runs low. The study employed a method combining machine learning and explainable artificial intelligence to analyze a dataset containing over 100,000 real travel records and simulate users' responses to related situations when the battery is extremely low.

The research results indicate that the immediate charging behavior after low battery levels is influenced by multiple interacting factors, including the battery state of charge (SOC) at the end of the trip, the total distance traveled, the time of day, and the frequency of charging. It is notable that evening trips and scenarios with low battery levels but longer distances significantly increase the likelihood of charging. By using XGBoost for powerful feature extraction and classification, the prediction model achieved a high accuracy rate (87%) and a high AUC score (0.90), confirming that the model can capture the real behavioral trends.

The integration of SHAP also provides transparent explanations for the model's output. This enhanced interpretability helps reveal the hidden logic behind user behavior, identifying that drivers with lower battery charge levels, fewer recent charging sessions, and longer travel times are more likely to charge immediately. These insights not only validate the reliability of the prediction system but also offer opportunities for the planning of new energy charging station construction.

The main contributions of this study are threefold. Firstly, it has developed a data-based framework to describe the behavioral characteristics of low-power driving during the use of electric vehicles. Secondly, it demonstrates that the behavior perception prediction model can achieve high accuracy and transparency, thereby establishing a connection between predictive analysis and decision support. Finally, this study showcases the potential of interpretable models in guiding the deployment of intelligent charging infrastructure and in the research of electric vehicle policies. These contributions are of practical significance for vehicle research and manufacturing enterprises, urban transportation planners, and intelligent energy systems.

5.2 Future Works

Although the current research has achieved encouraging results, it must be acknowledged that there are still some limitations. The model has not yet incorporated factors such as space, environment, or external policies that may affect charging behavior. Additionally, no sufficient research has been conducted on the temporal dynamic changes in behavior evolution over multiple days or usage cycles.

Integration of geographic data and infrastructure data: Incorporating elements such as geographical location information, the availability of nearby charging stations, and regional charging costs into the consideration scope will help address issues in the decision-making process more comprehensively.

Incorporation of environmental factors: External factors such as temperature, traffic congestion conditions, and weather conditions have been proven to affect battery consumption and the urgency of charging. If these factors are taken into account, they will be of great value.

Time-based behavior modeling: Using time-aware recurrent models or architectures can capture users' charging habits and deviations over longer periods, thereby enhancing the degree of personalization.

Adaptive decision support systems: Embedding the prediction engine into the on-board real-time system can generate proactive charging reminders, route re-planning, or dynamic battery charging status warnings based on driver behavior. 5. Policy optimization and simulation: The model can perform infrastructure planning based on simulations, which is helpful for identifying underserved areas, predicting demand hotspots, and evaluating the response effects of different charging incentive measures.

In summary, this research lays the foundation for the design of behavior-based electric vehicle systems. The predictive model is built based on actual usage behavior and demonstrates significant potential for contributions in intelligent electric vehicle management, user-centered travel services, and urban sustainable development strategies.

REFERENCES

Zhao, Y., Li, M., & Wang, H. (2022). *Charging-Related State Prediction for Electric Vehicles Using the Deep Learning Model*. Journal of Sustainable Transportation, 12(3), 45–67.

Li, X., Zhang, R., & Chen, Y. (2021). *Range Anxiety and Charging Behavior of Electric Vehicle Drivers: A Data-Driven Analysis*. Transportation Research Part C: Emerging Technologies, 123, 103123.

Zhang, W., Liu, T., & Wang, J. (2020). *Impact of Low Battery State-of-Charge on Electric Vehicle Driving Patterns*. Applied Energy, 278, 115532.

Liu, S., Wang, Y., & Huang, Z. (2023). *Data-Driven Characterization of Electric Vehicle Charging Behavior in Urban Areas*. IEEE Transactions on Intelligent Transportation Systems, 24(5), 4567–4578.

Wang, H., Chen, L., & Zhao, K. (2022). *Temporal and Spatial Patterns of Electric Vehicle Usage: Insights from Real-World Data*. Sustainable Cities and Society, 85, 103789.

Chen, J., Guo, X., & Li, Y. (2021). *Predicting Electric Vehicle Charging Demand Using Machine Learning: A Comparative Study*. Energy and Buildings, 245, 110876.

Kim, D., Park, S., & Lee, H. (2020). *Machine Learning-Based Prediction of Charging Station Utilization for Electric Vehicles*. Journal of Cleaner Production, 265, 123456.

Lundberg, S. M., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. Advances in Neural Information Processing Systems, 30, 4765–4774.

Molnar, C. (2020). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Open Access Book.

Guo, F., Yang, H., & Wang, X. (2021). *Optimal Charging Station Placement for Electric Vehicles: A Data-Driven Approach*. Transportation Research Part E: Logistics and Transportation Review, 152, 102345.

European Environment Agency (EEA). (2022). *Electric Vehicle Infrastructure and Policy Implications for a Sustainable Transport System*.

Appendix A Sample entries from the EV trip dataset

Contains representative rows of the structured dataset used for model training and evaluation. Fields include:

trip_start_time: Timestamp when the trip began

trip_end_time: Timestamp when the trip ended

soc_end: Final state-of-charge percentage at trip end

trip_distance: Distance driven in kilometers

next_charge_time: Time when the next charging session began

charged_immediately: Binary label indicating if charging started promptly (1) or not(0)

Appendix B Hyperparameter configuration for XGBoost

The XGBoost classifier was optimized using the following parameter values:

objective: ‘binary:logistic’

eval_metric: ‘auc’

eta: 0.1

max_depth: 6

subsample: 0.8

colsample_bytree: 0.8

num_boost_round: 100

early_stopping_rounds: 10

These parameters were selected based on empirical testing to balance accuracy and generalization.

LIST OF PUBLICATIONS

Journal Articles

Qasem, S. N., Shamsuddin, S. M., Hashim, S. Z. M., Darus, M., & AlShammari, E. (2013). Memetic multiobjective particle swarm optimization based radial basis function network for classification problems. Information Sciences, 239, 165–190. https://doi.org/10.1016/j.ins.2013.03.021. (Q1, IF: 4.305)

Qasem, S. N., & Shamsuddin, S. M. (2011). Radial basis function network based on time variant multi-objective particle swarm optimization for medical diseases diagnosis. Applied Soft Computing, 11(1), 1427–1438. https://doi.org/10.1016/j.asoc.2010.04.014. (Q1, IF:3.907)

Shen, L. W., Asmuni, H., & Weng, F. C. (2015). A modified migrating bird optimization for university course timetabling problem. Jurnal Teknologi, 72(1), 89–96. https://doi.org/10.11113/jt.v72.2949. (Indexed by SCOPUS)

Conference Proceedings

Muhamad, W. Z. A. W., Jamaludin, K. R., Ramlie, F., Harudin, N., & Jaafar, N. N. (2017). Criteria selection for MBA programme based on the mahalanobis Taguchi system and the Kanri Distance Calculator. In 2017 IEEE 15th Student Conference on Research and Development (SCOReD) (pp. 220–223). IEEE. https://doi.org/10.1109/SCORED.2017.8305390. (Indexed by SCOPUS).