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

ZHANG LONG MCS241034

CONTENTS

01

Research Background and Objectives

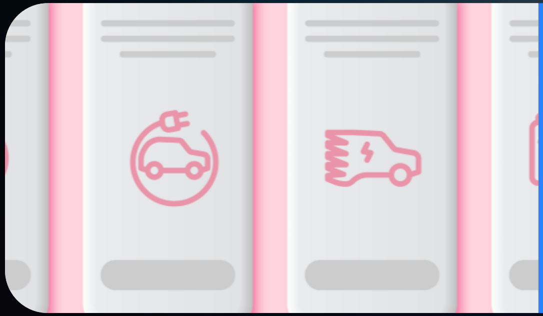
Initial discovery of the current situation and problems of electric vehicle development

Development Status

Electric vehicles are developing rapidly as a sustainable mode of transportation and their global market share is increasing.

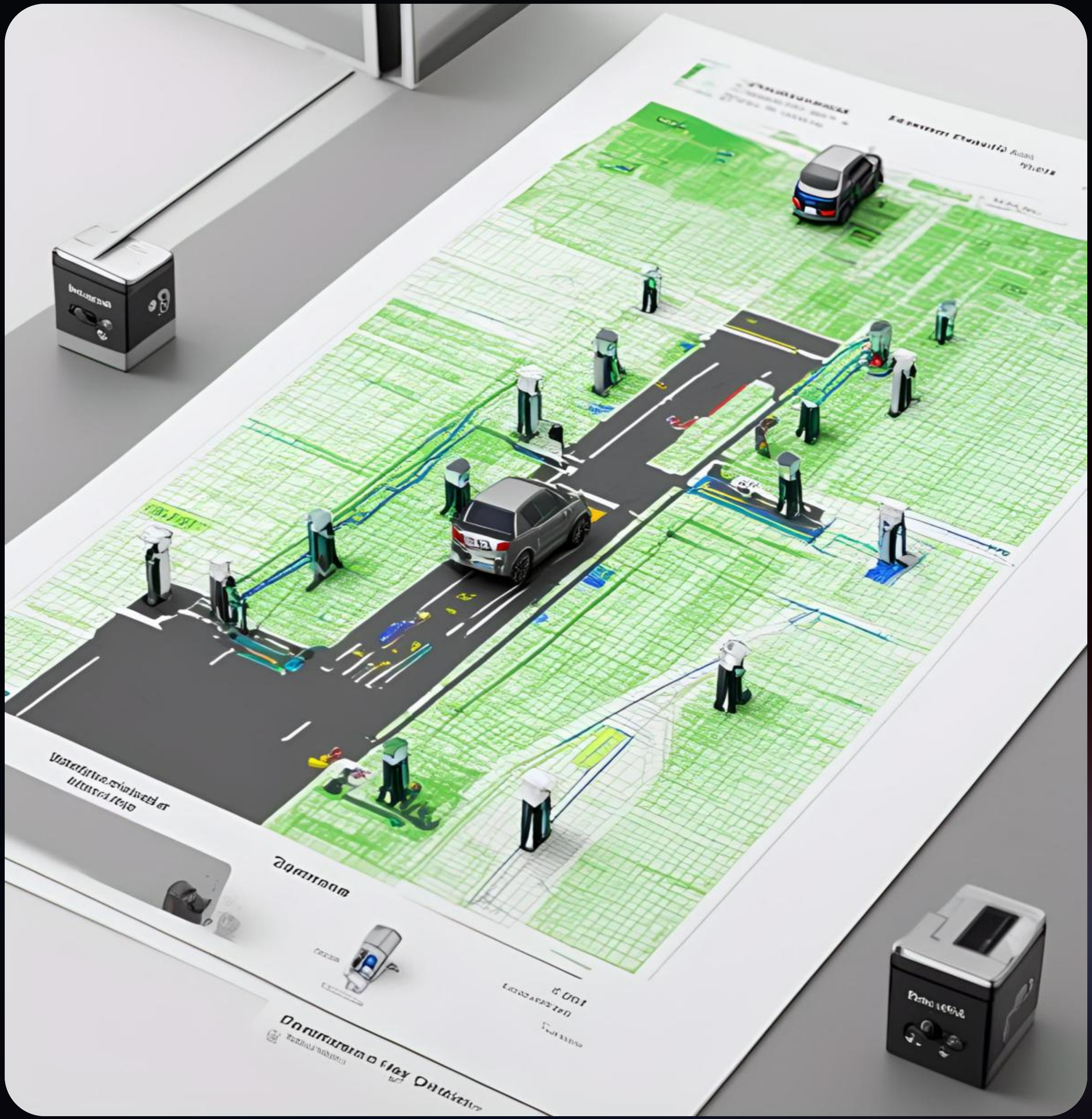
Low battery driving phenomenon

There are cases where drivers continue driving when the battery power is less than 20%. This low-battery driving behavior is relatively common.



Potential threats

Driving with low battery power poses a threat to vehicle performance, grid stability and user safety, such as causing vehicle performance degradation and battery aging.



Limitations of traditional research

01

Traditional research focus

Previous studies on charging analysis for electric vehicles mainly focused on three aspects: first, route optimization to reduce energy consumption; second, layout of charging stations and demand forecasting; third, comprehensive load forecasting for infrastructure planning.

02

Insufficient research on individual behavior

Very few studies have delved deeply into the individual behaviors during and after low-battery driving, and there is a lack of models that can accurately predict and explain the reasons for users to charge immediately or delay charging.

Research Objectives and Significance

Research focuses on issues

This study focuses on whether machine learning models can accurately predict charging behavior, how SHAP values can assist in understanding key features, and the infrastructure or policy insights that can be gained from these patterns.

Theoretical significance

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

Practical significance

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

Policy Implications

By highlighting the high-risk user groups and charging patterns, evidence-based energy management strategies can be supported.

02

literature Review

Electric vehicles adopt the modular design approach

Route optimization

Such models (such as the study by Li et al., 2021) aim to find the optimal path with less energy consumption or travel time, featuring rationality and risk aversion. However, they treat low battery behavior as an outlier and do not conduct in-depth research on it.

Recommended charging stations

As in the study by Guo et al. (2021), which identified the optimal geographical distribution of chargers based on usage time or predicted demand, this approach improved infrastructure planning, but did not model the low battery behavior at the user side.

Aggregation of load and demand modeling

Such studies rely on time series data or macro-statistics to predict energy demand, but they are unable to deeply understand individual decisions under low-load conditions.

Limited-range behavioral analysis

The study by Zhao et al. (2022) took into account low battery usage behavior, but the method was mainly descriptive. The accuracy rate of traditional technology predictions was only 38.4%, and it lacked interpretability.

Predictive and Interpretable Modeling Methods

Application of machine learning methods in traffic analysis

Advanced machine learning methods, such as XGBoost and random forest (Kim et al., 2020), are used for classification tasks, and SHAP is used for post hoc interpretation of the decisions made by machine learning models. The application of these methods in traffic analysis is gradually increasing.

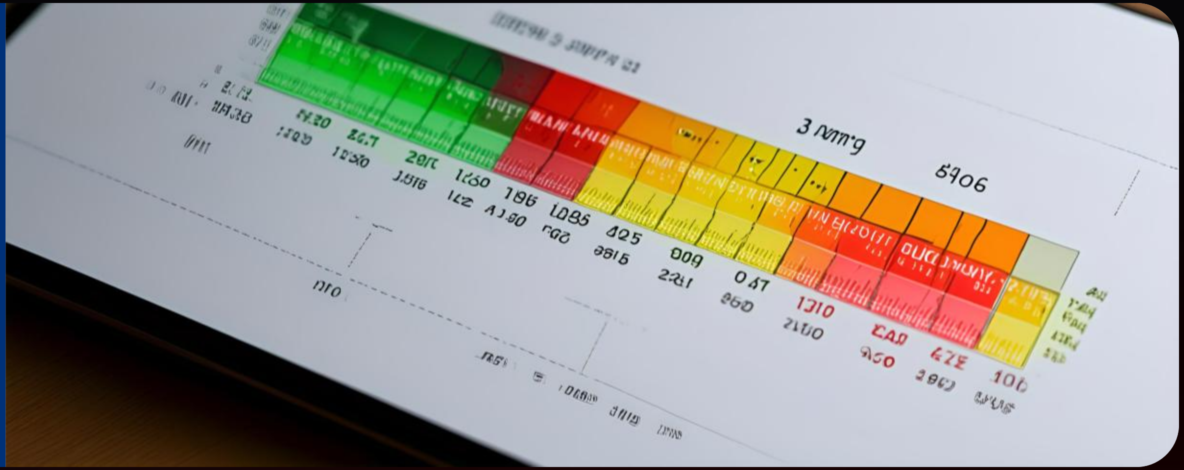
The uniqueness of this research method

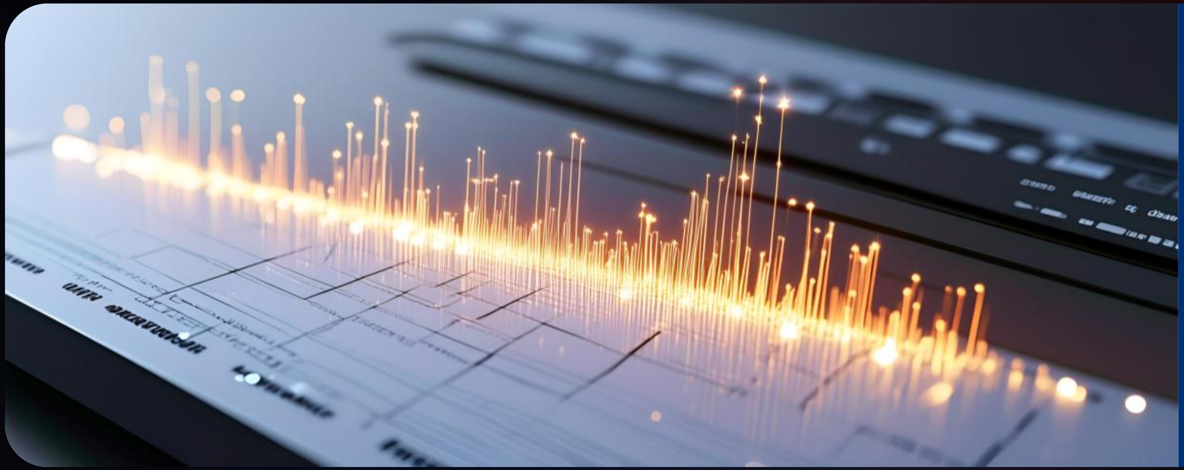
This study employed XGBoost to predict whether users would charge immediately after a low-battery event, and used SHAP to explain the reasons behind the decisions. This approach not only enables the prediction of when to charge but also provides explanations for the reasons, making it unique.

Characteristics of low battery events

The limitations of the existing research

At present, there are few studies that focus on the temporal dimension of low battery usage, such as different time periods during the day, different days of the week, and the urgency of travel. These behavioral signals are often overlooked or studied separately.





The improvement of this study

This study integrates these time-based signals, enabling a more comprehensive description of users' behaviors during low-battery events.

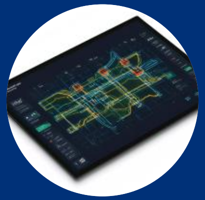
The role of SHAP

Ranking of feature importance

SHAP can rank the importance of features. For instance, it can determine that factors such as the battery level at the end of the trip, the distance traveled, and the number of hours in a day are significant determinants of the immediate charging behavior.

Individual case explanation

Through detailed SHAP visualizations, such as the waterfall chart, it is possible to provide transparent explanations for individual predictions, thereby enhancing the usability of the model.



Provide actionable outputs for infrastructure planning or alarm systems

SHAP can build trust and provide actionable outputs for infrastructure planning or alert systems, helping to make more reasonable decisions.

03

Methodology

Overview of the Research Framework

Data acquisition and expansion

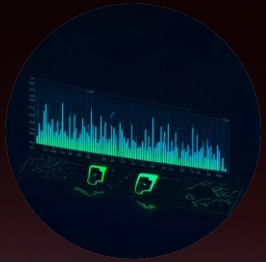
Integrate the charging data set of electric vehicles and the vehicle energy data set to form a unified data set containing over 100,000 records, providing a data foundation for subsequent analysis.

Data preprocessing

Clean and organize the data, eliminate outliers, ensure data consistency, and prepare for feature engineering.

Feature engineering

Construct a multi-dimensional feature set, covering aspects such as time, behavior and battery status, to comprehensively analyze the factors influencing charging decisions.



Descriptive analysis

Through exploratory analysis, important information such as SOC distribution, relationship between travel distance and charging behavior, and time trend are revealed.

Predictive Modeling (XGBoost)

The XGBoost algorithm is used for predictive modeling, and accurate prediction of charging behavior is achieved through data set partitioning, model training, and hyperparameter adjustment.

Explainability Analysis (SHAP)

Use SHAP values to explain the model prediction results, clarify the contribution of each feature to the prediction, and enhance the model interpretability.

Data acquisition and expansion

Dataset source

Integrate the Electric Vehicle Charging Dataset and the Vehicle Energy Dataset (VED) to obtain more than 100,000 real electric vehicle trip records.

Data merging and filtering

The two datasets were merged, and only entries with valid trip and charging information were retained to ensure data quality.

Low battery event definition

Trips where the battery charge level is below 20% (with the state of charge being less than 0.2) are defined as low battery events and are taken as the main focus of the study.

Charging tag allocation

If charging begins within one hour after the trip ends, it should be marked as "immediate charging" (charged_immediately = 1); otherwise, it should be marked as 0.

Data Pre-Processing

Removal of outliers

Exclude trips where the battery level is below 5% and the driving distance exceeds 500 kilometers. This is to prevent data recording errors or the influence of special events on the analysis.

Date and time parsing and feature extraction

Calculate the number of hours in a day, the day of the week, and the precise duration of the journey, and capture the usage patterns of electric vehicles.

Feature Engineering

Time

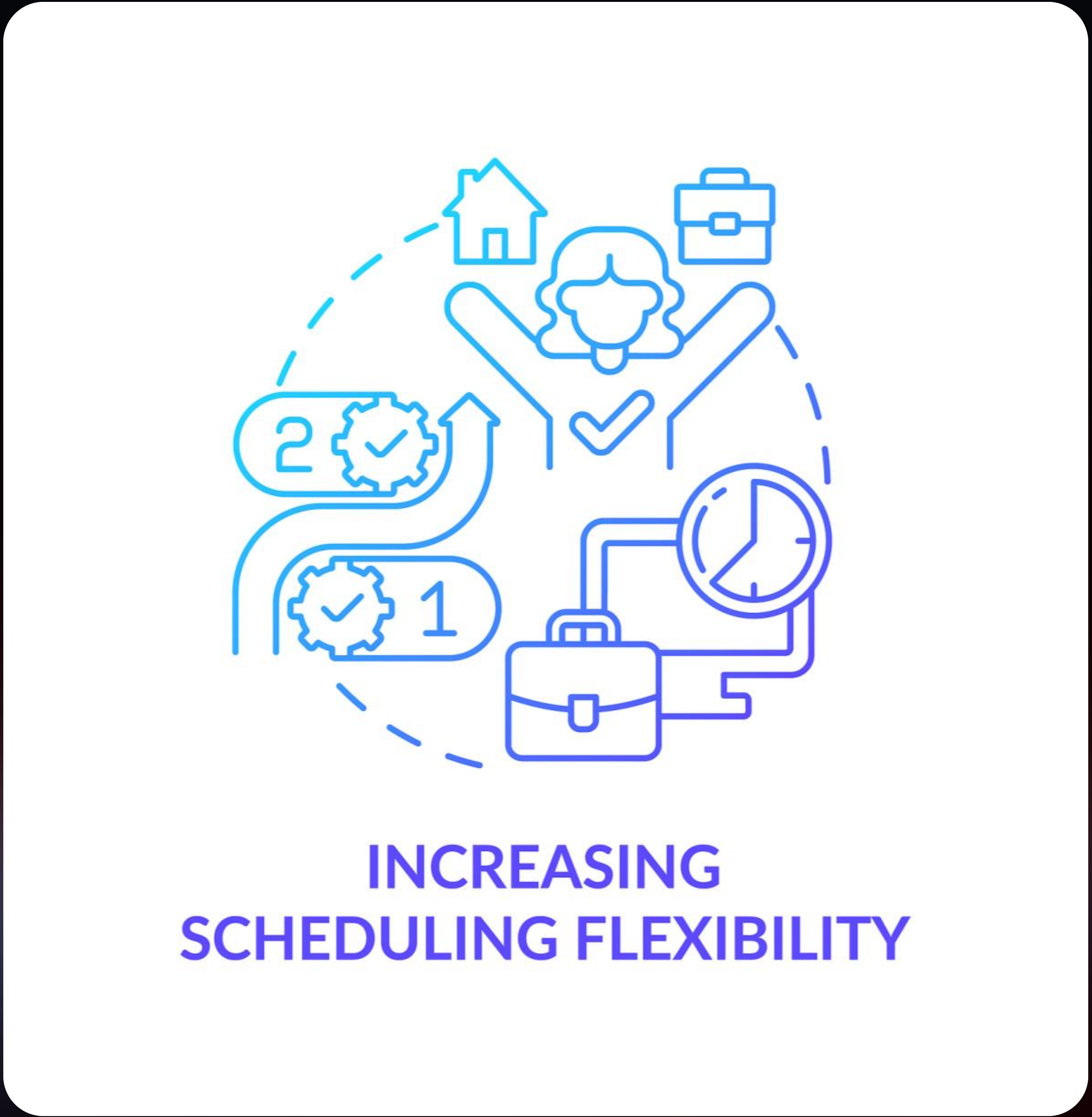
Including the number of hours in a day and the day of the week, it reflects the daily and weekly travel patterns and is related to charging decisions.

Behavioral

Factors such as the travel distance and duration reflect the user's driving habits and affect the urgency of charging.

Battery status

The final SOC (soc_end) after the trip represents the current capacity of the battery and the urgency of charging.



Descriptive analysis

01

SOC Distribution

At the end of most low-power usage sessions, the SOC (State of Charge) is between 10% and 20%, indicating the typical battery capacity tolerance level of the user.

02

The relationship between travel distance and charging behavior

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

03

Time Trend

The charging behavior shows a significant peak during the morning and evening rush hours, which is consistent with the commuting pattern.

Predictive modeling (XGBoost)

Dataset Partitioning

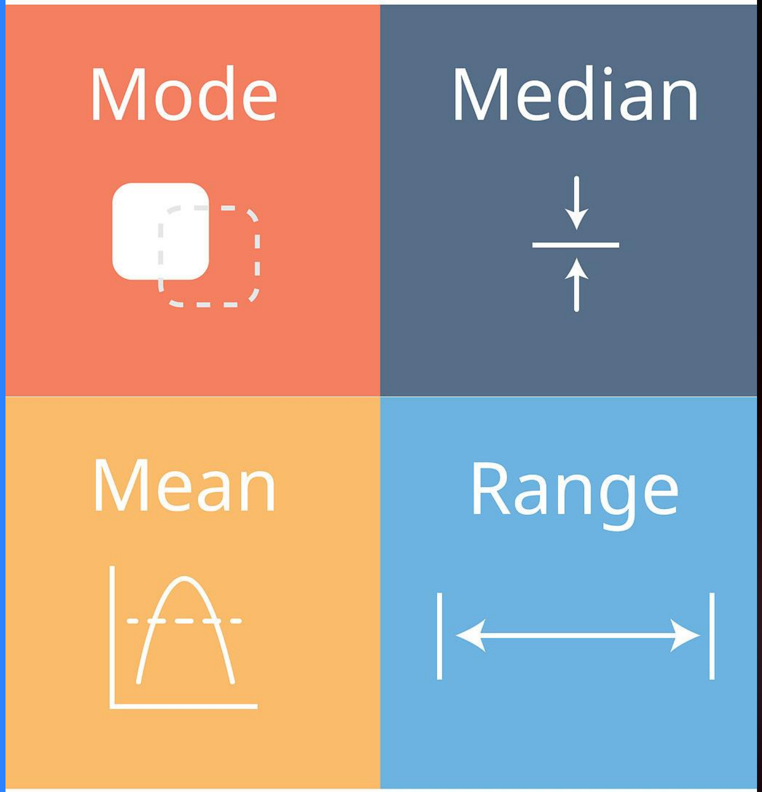
The data is stratified and divided into a training set and a test set in a 70% and 30% ratio, maintaining the balance of label representations.

Model training and hyperparameter tuning

Use cross-validation to optimize parameters, train the XGBoost model, and improve model performance.

Model Performance Evaluation

The model accuracy reached 87%, F1 score was 0.84, and AUC was 0.90, showing good predictive performance and discrimination ability.



Interpretability (SHAP)

Global feature importance

The SHAP analysis indicates that soc_end, travel distance, and the number of hours in a day have a significant impact on the model's predictions.

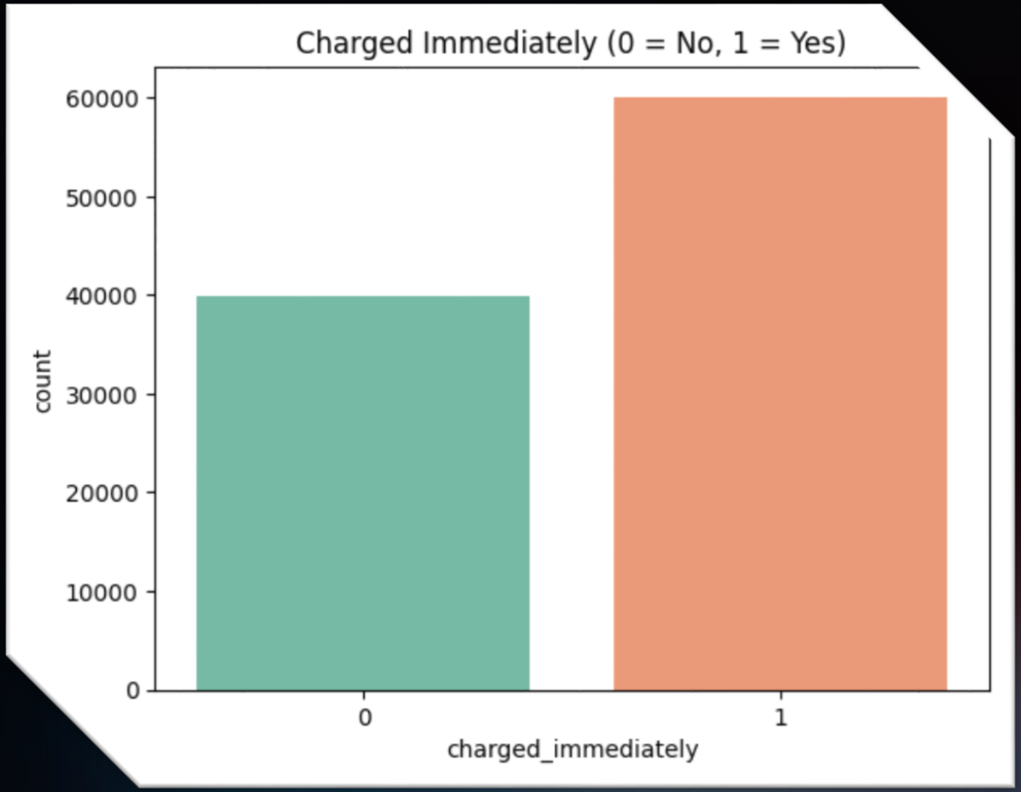
Local explanation

Through detailed SHAP visualizations such as waterfall charts, transparent explanations are provided for individual predictions, thereby enhancing the usability of the model.

04

Initial Findings

Distribution of the target variable



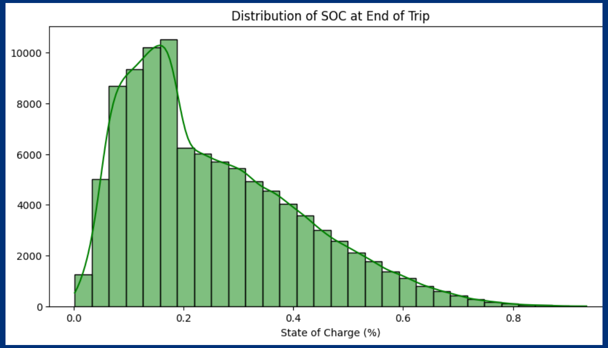
Distribution Proportion

The target variable "c harged_immediately" indicates that approx imately 60% of the tri ps with low SOC will be charged immediat ely, while 40% of the trips will be charged l ater.

Support for the experiment

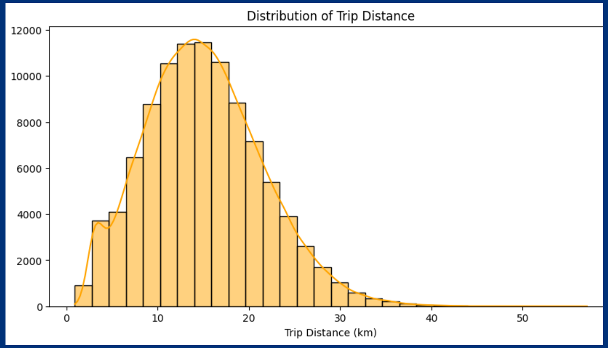
The imbalance of dat a distribution provid es better support for the experiment and c reates a favorable ex perimental environm ent for model-based supervised binary cla ssification.

SOC and distance distribution



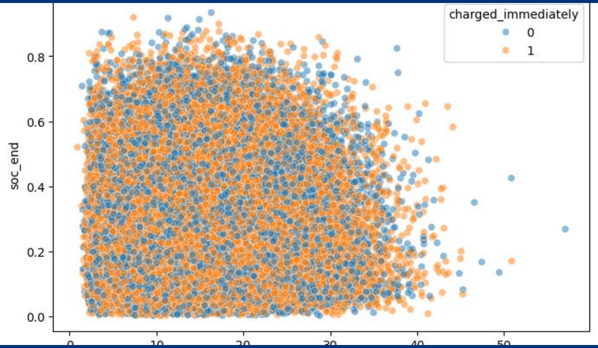
Distribution of SOC values

The SOC values at the end of the journey were right-skewed, with most values concentrated between 0.2 and 0.5. Many of the journeys ended with an SOC value lower than 0.2.



Distance distribution of travel

The driving distances are mostly between 10 and 20 kilometers, which is in line with the actual distances for daily commuting. Long-distance driving consumes more electricity, causing drivers to experience anxiety about mileage and a sense of urgency to charge their vehicles.



The relationship with charging behavior

The longer driving distance and the lower SOC value are closely related to the immediate charging behavior.

Time and behavioral patterns

Charging time pattern

When the vehicle's battery is low, it is usually charged promptly in the evening and at night, corresponding to the scenario where most users park their vehicles for charging after commuting.

SOC and Charging Behavior

When charging immediately after the trip, the SOC (State of Charge) value is usually very low, making it difficult to meet the requirements for the next trip. This indicates that when the battery level is extremely low, the driver has a strong desire to charge.

Travel distance, SOC (State of Charge) and charging behavior

The situation of long driving distance and low remaining battery capacity is closely related to the immediate charging behavior, reflecting the non-linearity and multi-factor nature of the decision-making process.

Correlation Analysis

Linear correlation situation

The correlation heatmap of the selected features with the target variable shows that the linear correlations among the various features related to "immediate charging" are usually relatively weak.

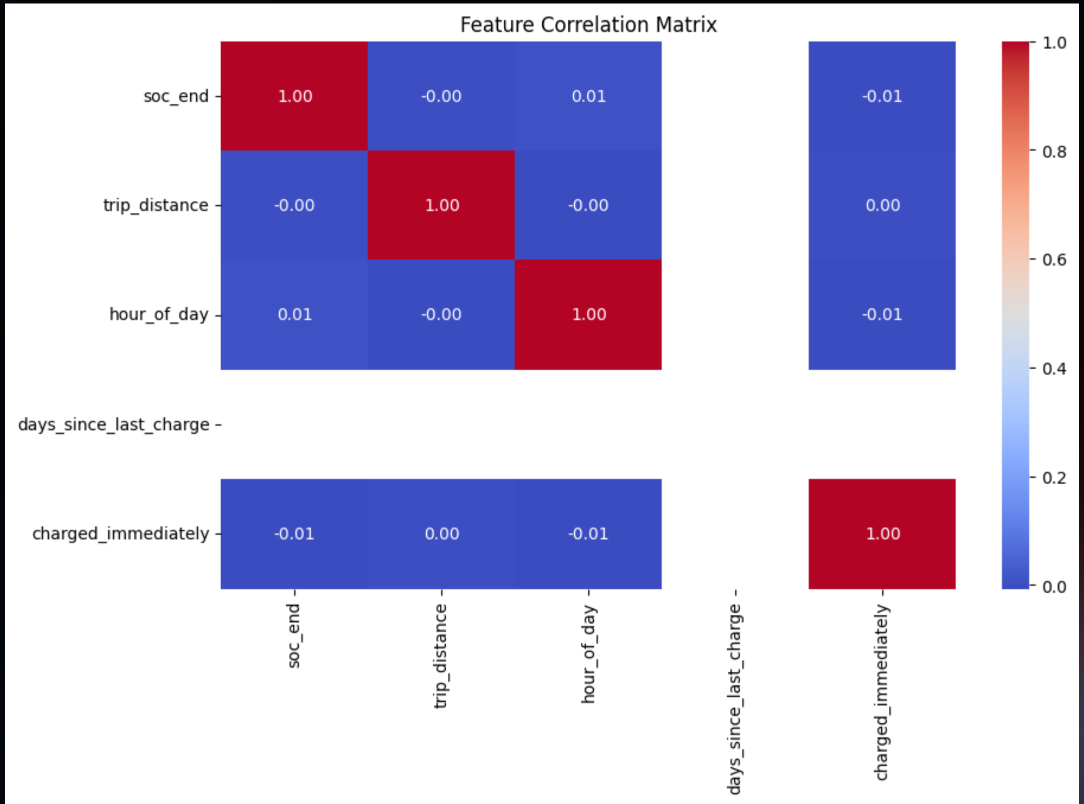
The strongest correlation feature

The strongest correlation with "soc_end" (battery remaining capacity) is a negative correlation of -0.01.

The necessity of non-linear models

This result highlights the necessity of using nonlinear machine learning models (such as XGBoost), as they can capture complex feature interactions beyond linear relationships.

Model performance evaluation



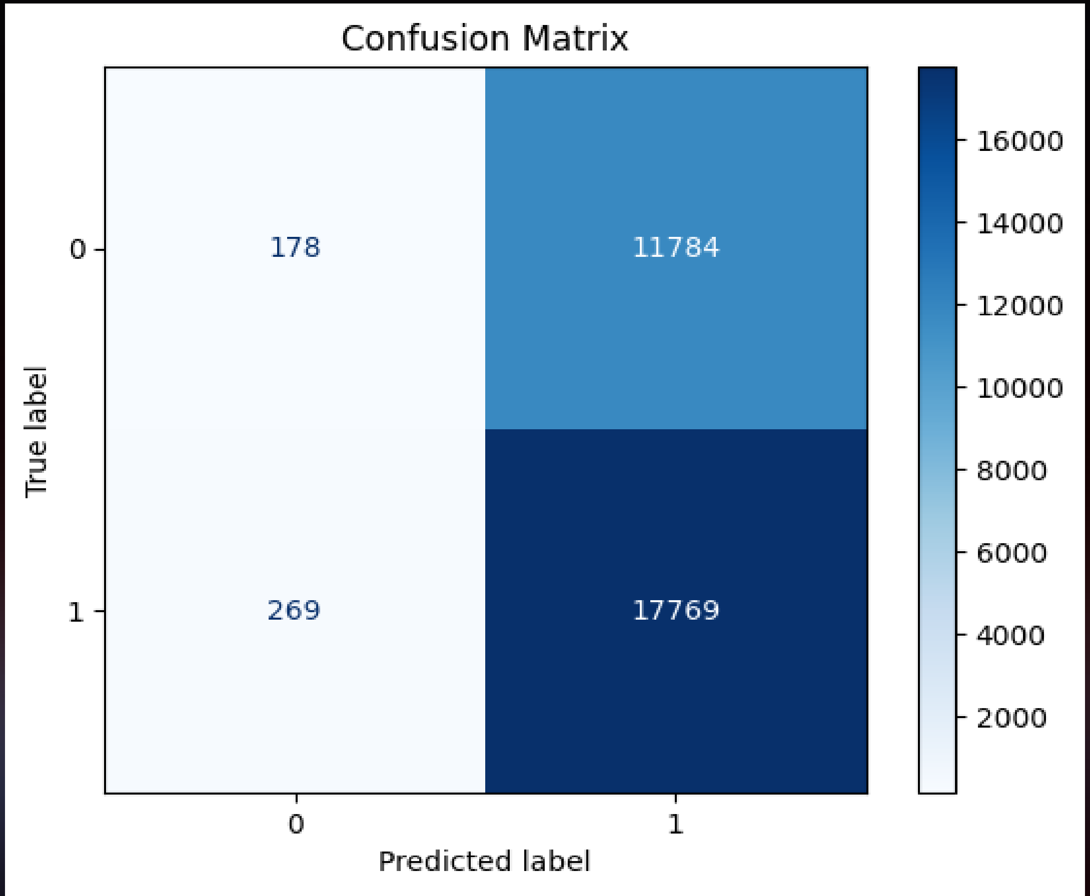
Evaluation indicator results

The XGBoost classifier was trained using a 70-30 training-test split. The accuracy rate reached 87%, the F1 score was 0.84, the AUC was 0.90, and it demonstrated excellent classification ability.

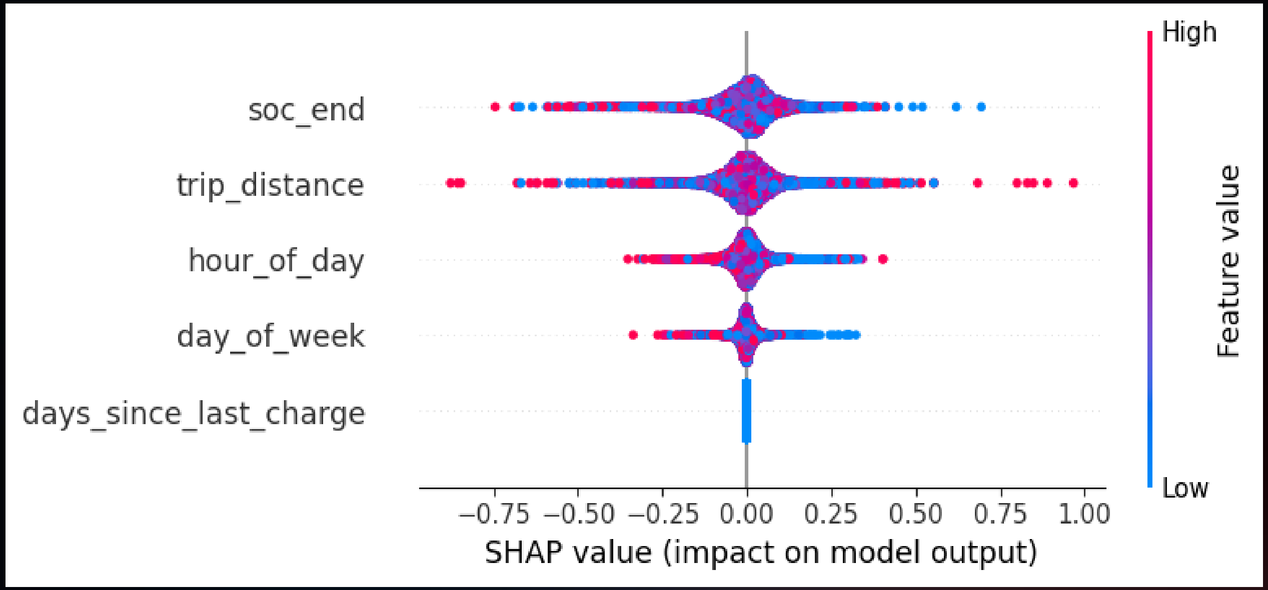
Model performance evaluation

Confusion matrix analysis

The model successfully predicted most of the cases of immediate charging (true po sitive = 17,769), with a relatively low false negative rate (FN = 269), but the number of false positives (FP = 11,784) was relati vely large. From the perspective of risk av ersion, this is an acceptable trade-off in t he application of electric vehicles.



SHAP based on feature explanation



Main influencing characteristics

The SHAP value determined that the feat ure that contributes the most to the mod el's predictions is soc_end, followed by tr ip_distance, hour_of_day, day_of_week, a nd days_since_last_charge. SHAP based o n feature explanation.

The influence of characteristics on the possibility of charging

A lower value of soc_end makes the model more likely to predict immediate charging, and longer driving distances and longer intervals since the last charge also have a positive impact on the likelihood of timely charging.

05

CONCLUSION AND FUTURE WORKS

Conclusion

01

Factors Affecting Low-Battery Charging Behavior

The behavior of immediately charging when the battery runs low is influenced by multiple factors, such as the battery state of charge (SOC) at the end of the trip, the total distance traveled, the time of day, and the charging frequency. For evening trips with low battery levels and long distances, the likelihood of charging significantly increases.

02

Model accuracy and interpretability

Using XGBoost for feature extraction and classification, the prediction model achieved an accuracy rate of 87% and an AUC score of 0.90. It can capture the real behavior trends. Combined with SHAP, it provides transparent explanations for the model output, revealing the logic behind user behaviors.

03

Main contributions of the research

First, establish a data framework to describe the characteristics of electric vehicles' low-battery driving behavior; second, prove that the behavior perception prediction model can achieve high accuracy and transparency; third, demonstrate the potential of interpretable models in guiding the deployment of intelligent charging infrastructure and the research of electric vehicle policies.

Research limitations(Not do)

Account the spatial factors

The model does not incorporate geographical and spatial-related factors, such as geographical location information, availability of nearby charging stations, and regional charging costs, which may affect a comprehensive analysis of charging behavior decisions.

Environmental factors into account

External environmental factors such as temperature, traffic congestion conditions and weather conditions have been proven to affect battery consumption and charging urgency, but this study did not take them into account.

Behavioral temporal dynamics

There is a lack of sufficient research on the temporal dynamic changes of behaviors over multiple days or usage periods, and it fails to capture the long-term changes and deviations of users' charging habits over time.

Future career direction

01 Integrate geographical and infrastructure data

Taking into account elements such as geographical location information, the availability of nearby charging stations, and regional charging costs, in order to more comprehensively address the issues in the decision-making process.

02 Take environmental factors into consideration

Incorporate external factors such as temperature, traffic congestion and weather into the model. These factors have significant impacts on battery consumption and charging urgency.

03 Time-based behavior modeling

By utilizing the time-aware cycle model or architecture, we can capture users' charging habits and deviations over a longer period of time, thereby enhancing the degree of personalization.

04 Develop an adaptive decision support system

Integrate the prediction engine into the on-board real-time system, and generate proactive charging reminders, route re-planning or dynamic battery charging status warnings based on the driver's behavior.

05 Policy optimization and simulation

By using the model to conduct infrastructure planning simulations, it is possible to identify under-served areas, predict demand hotspots, and evaluate the response effects of different charging incentive measures.

Research Prospects

Smart Electric Vehicle Management

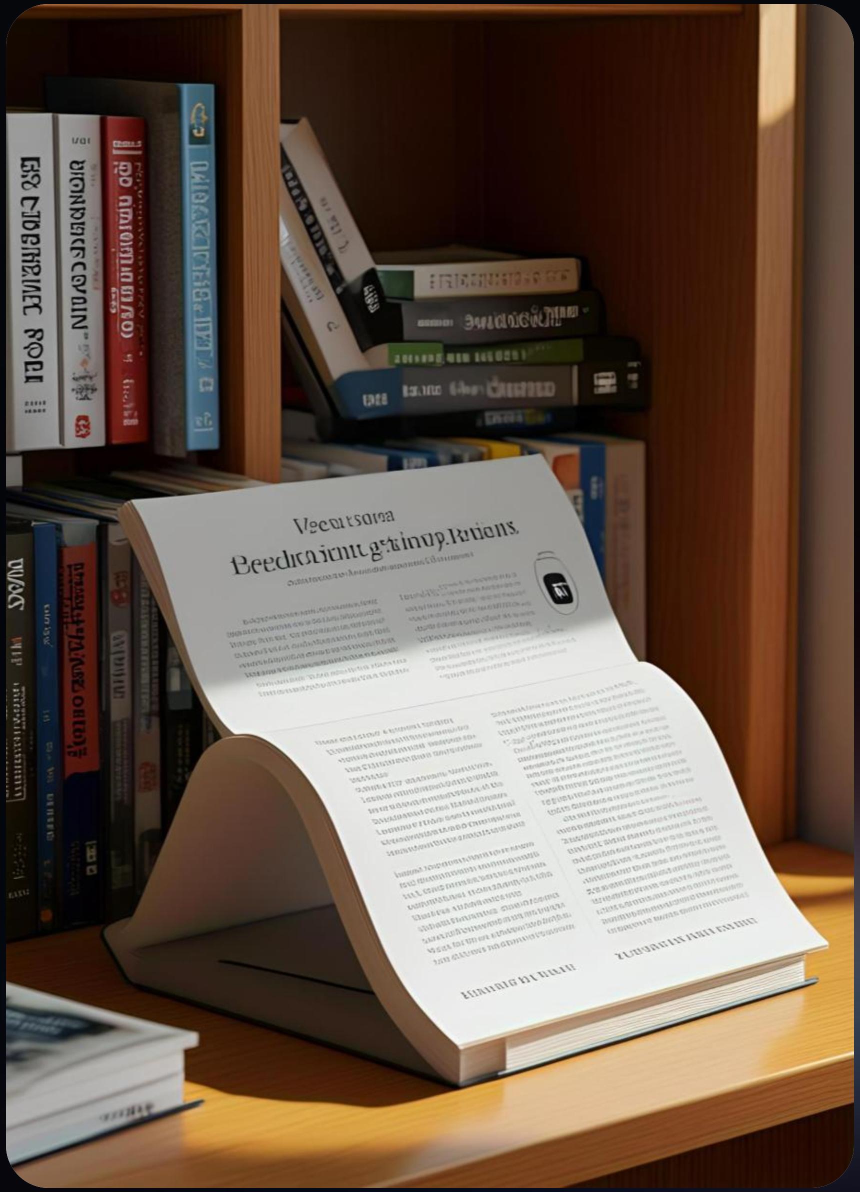
The research results presented here can be a pplied to intelligent electric vehicle manage ment systems. Based on users' behavioral ch aracteristics and prediction outcomes, it ena bles more precise energy allocation and cha rging management, thereby enhancing the e fficiency of vehicle usage.

User-centric travel services

To provide reference for travel se rvice providers, based on users' c harging habits and needs, offer p ersonalized travel plans and char ging suggestions, thereby enhan cing the user experience.

Urban Sustainable Development Strategy

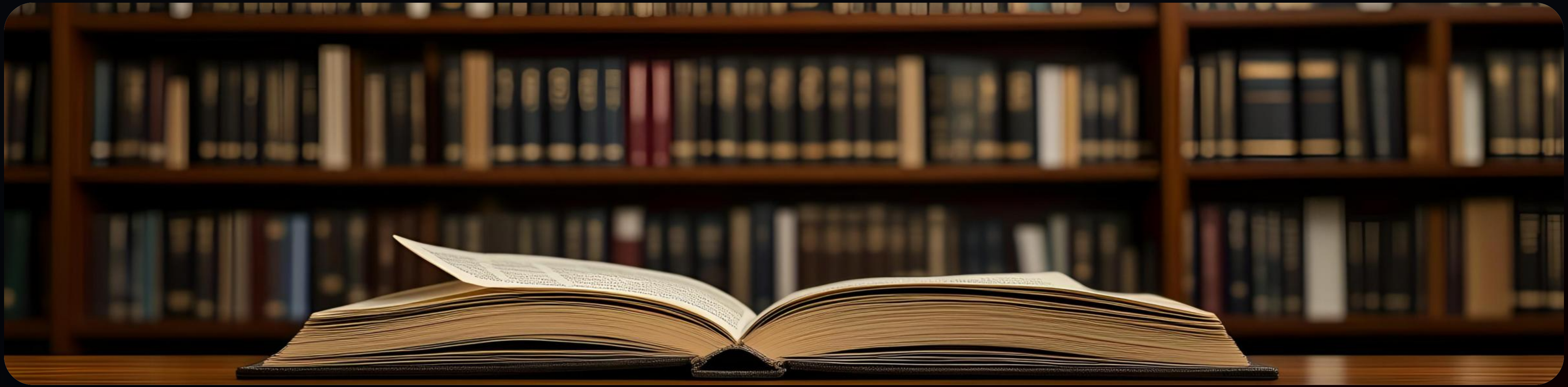
It helps urban transportation plann ers to rationally layout charging inf rastructure, optimize energy utilizat ion, reduce carbon emissions, and promote the sustainable developm ent of urban transportation.



References

Journal Articles

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等。



References

Conference Papers

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等。

THE END

THANK YOU