

An Explainable Machine Learning Method for Predicting Electric Vehicle Prices in the Iraqi Market

Dr. Hassan Kassem¹, Dr. Tariq Yousef Bishtawi^{*2}

¹ Department of engineering & computing Technology | Arab University College of Technology | Jordan

² Department of Computer Science | Faculty of Information Technology | Amman Arab University | Jordan

Received:

01/03/2026

Revised:

15/03/2026

Accepted:

12/04/2026

Published:

15/06/2026

* Corresponding author:

t.bishtawi@aaau.edu.jo

Citation: Kassem, H., & Bishtawi, T. Y. (2026). An explainable machine learning method for predicting electric vehicle prices in the Iraqi market. *Journal of Engineering Sciences and Information Technology*, 10(2), 118 – 135.

<https://doi.org/10.26389/AJSRP.T030326>

2026 © AISRP • Arab Institute for Sciences & Research Publishing (AISRP), United States, all rights reserved.

• Open Access



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) [license](https://creativecommons.org/licenses/by-nc/4.0/)

Abstract: The emerging market of electric vehicles (EV) in the emerging economies, including in Iraq, is gradually turning into a significant issue of obscure valuation because the conventional internal combustion engine (ICE)-based model of price formation is not appropriate in determining and describing the special determinants of EV price formation. It is against this backdrop that the present study is an endeavor to introduce an explainable machine learning (ML) model which is a precise predictor of prices of used EVs in Iraq, to be more precise, the capital of Baghdad. The methodological strategy presupposed the systematic gathering of much volume of EV data, including such variables as battery capacity, driving range, and a package of local market variables. A regional database of 1420 electric vehicle advertisements on the Iraq market were gathered and assessed. CatBoost, XGBoost, and Random Forest tree-based regression algorithms were utilized to make predictions, and SHapley Additive exPlanations (SHAP) were added to increase transparency and interpretability. The results of the empirical analysis proved that the CatBoost model showed the best predictive performance, with the R^2 equal to 0.94, Mean Absolute Error (MAE) equal to 1,245,000 IQD, and Root Mean Squared Error (RMSE) equal to 2,180,000 IQD. Moreover, the findings demonstrate that EV specific attributes, especially battery capacity and driving range are the most influential price factors, which outweigh traditional variables. The supremacy of these characteristics is supported by SHAP analysis which shows these characteristics are supreme. The best-performing model that was tested was CatBoost and it performed particularly well with high-cardinality categorical predictors. Overall, this explicable ML format can improve price transparency and allow informed decision-making among the consumers and dealerships of Iraq and policy makers, which in its turn, has a positive impact on the development of trust and promotes sustainable EV adoption on a regional level.

Keywords: explainable machine learning; SHAP; battery capacity; Iraqi automotive market; electric vehicle price prediction; used electric vehicles.

طريقة تعلم آلي قابلة للتفسير للتنبؤ بأسعار السيارات الكهربائية في السوق العراقية

الدكتور/ حسن قاسم¹، الدكتور/ طارق يوسف البشتاوي^{*2}

¹ قسم الهندسة وتكنولوجيا المعلومات | الكلية الجامعية العربية للتكنولوجيا | الأردن

² قسم علم الحاسوب | كلية تكنولوجيا المعلومات | جامعة عمان العربية | الأردن

المستخلص: يشهد سوق السيارات الكهربائية الناشئ في الاقتصادات الناشئة، بما فيها العراق، تحولاً تدريجياً نحو مشكلة تقييم غامضة، نظراً لعدم ملائمة نموذج تحديد الأسعار التقليدي القائم على محركات الاحتراق الداخلي في تحديد ووصف العوامل المحددة لأسعار السيارات الكهربائية. وانطلاقاً من هذا، تسعى هذه الدراسة إلى تقديم نموذج تعلم آلي قابل للتفسير، يُعدّ مؤشراً دقيقاً لأسعار السيارات الكهربائية المستعملة في العراق، وتحديدًا في العاصمة بغداد. وقد اعتمدت الاستراتيجية المنهجية على جمع كميات كبيرة من بيانات السيارات الكهربائية بشكل منهجي، بما في ذلك متغيرات مثل سعة البطارية، ومدى القيادة، ومجموعة من متغيرات السوق المحلية. وتم جمع قاعدة بيانات إقليمية تضم 1420 إعلاناً عن سيارات كهربائية في السوق العراقية، ثم جرى تقييمها. واستُخدمت خوارزميات الانحدار القائمة على أشجار CatBoost وXGBoost وRandom Forest للتنبؤ، كما أُضيفت تقنية SHapley Additive ExPlanations (SHAP) لزيادة الشفافية وقابلية التفسير. أثبتت نتائج التحليل التجريبي أن نموذج CatBoost أظهر أفضل أداء تنبؤي، حيث بلغ معامل التحديد R^2 0.94، ومتوسط الخطأ المطلق (MAE) 1,245,000 دينار عراقي، وجذر متوسط مربع الخطأ (RMSE) 2,180,000 دينار عراقي. علاوة على ذلك، تُشير النتائج إلى أن خصائص السيارات الكهربائية، وخاصة سعة البطارية ومدى القيادة، هي أكثر عوامل السعر تأثيراً، متفوقةً على المتغيرات التقليدية. ويؤكد تحليل SHAP هذه الخصائص، مُظهرًا أهميتها القصوى. وكان نموذج CatBoost هو الأفضل أداءً بين النماذج التي تم اختبارها، وقد حقق أداءً متميزاً، لا سيما مع المتغيرات التصنيفية ذات العدد الكبير من المتغيرات. وبشكل عام، يُمكن لهذا النموذج القابل للتفسير في مجال التعلم الآلي تحسين شفافية الأسعار، وتمكين المستهلكين وتجار السيارات في العراق وصناع السياسات من اتخاذ قرارات مستنيرة، مما يُسهم بدوره في تعزيز الثقة ودعم تبني السيارات الكهربائية بشكل مستدام على المستوى الإقليمي.

الكلمات المفتاحية: التعلم الآلي القابل للتفسير؛ SHAP؛ سعة البطارية؛ سوق السيارات العراقي؛ توقع أسعار السيارات الكهربائية؛ السيارات الكهربائية المستعملة.

1. INTRODUCTION

Currently, the automotive sector in the world is undergoing a radical change, driven by the growing environmental concerns, the rapid technological advancements, and the policy frameworks that are actively promoting the use of electric vehicles (EVs). The paradigm shift is not exclusive to the developed economies but is slowly trickling over to the emerging economies, such as Iraq. Traditionally, Iraq which is an oil-producing country has relied on internal combustion engines (ICE) cars but there is a growing appreciation of the potential values of EVs whether through emission reductions or the reduction in operating expenses of the average consumer. The appearance of a second-hand EV market, especially in the capital of Baghdad city in Iraq, is unique in terms of its challenges and opportunities in relation to pricing transparency, market efficiency, and consumer decision-making.

In contrast to the already established depreciation curves and measures of valuation that operate ICE vehicles, EVs present new price determinants that are critical, with battery capacity and driving range being the key ones. These EV-specific features change the traditional car valuation paradigm in its core. Proper price forecasting of used EVs is thus critical to a group of stakeholders in the Iraqi market. To its users, price estimates that are reliable make the process of buying products more informed; it also provides just a way of countering information asymmetry that is rife in online markets. To dealerships and sellers, information-driven pricing strategies can maximize the revenue and inventory turnover. To policy makers and market researchers, the factors behind the valuation of EV may provide them an insight into the tendency of technology adoption and market maturation in Iraq. Although the amount of global research into the topic of EV price prediction using machine learning methods is growing, there is an evident lack of research interested in the Iraqi market, in particular. The current literature does not tend to provide the specific socio-economic, infrastructural, and regulatory subtleties that need to be taken into account in such countries as Iraq. The present paper aims to fill this gap by suggesting an explainable machine learning model in which the specifics of the Iraqi automotive market are considered.

1.1 Problem statement

The very rapid development of a secondary market in used electric vehicles (EVs) in Iraq (and in specific urban areas, like Baghdad) has demonstrated a major vulnerability in the market infrastructures: a lack of a standardized, transparent, and data-driven system of valuation. The EVs have new and prevailing value determinants, the first one being battery capacity and driving range, unlike the traditional internal combustion engine (ICE) vehicles, which have well-established depreciation models. Absence of a sound pricing methodology breeds a high level of information asymmetry, low consumer confidence, challenges dealership in management of inventories, and clouds market dynamics to policymakers. The issue that is tackled in this research is incorrect and transparent price discovery of used EVs in the emerging Iraqi market.

1.2 Research Objectives

The main aim of the study is to create and test an explainable machine learning model that would be able to forecast the prices of used electric vehicles in the Iraqi market with high accuracy.

The following are the secondary objectives:

- To produce a market-specific dataset of used EVs in Iraq by gathering information on online markets and local dealerships with emphasis on traditional and EV-specific features.
- To determine the most important variables of used EV prices in the Iraqi market, paying special attention to the role of battery capacity, driving range, and local market variables, including tariffs on imports and the availability of the brand.
- To compare the performance of different tree-based regression models (CatBoost, XGBoost, Random Forest) when predicting EV prices and to choose the best and the most reliable model.
- To improve the transparency and interpretability of the predictive model through the application of SHapley Additive exPlanations (SHAP), which can help offer both global and local information on the determinants of prices.

1.3 Research questions and Hypotheses.

The research questions and hypothesis in this study are as follows:

RQ1: What are the most influential determinants of used electric vehicle prices in the Iraqi market?

H1: In Iraq, battery capacity and driving range as EV-specific features will emerge as the most influential predictors in used EV prices, compared to the classical predictors that include brand, model, and mileage.

RQ2: To what extent are machine learning models applicable to predicting the prices of used EVs in the Iraqi market?

H2: Tree-based ensemble models, specifically CatBoost, will be highly predictive (in terms of R^2 and low error values) when trained on a localized Iraqi EV dataset.

RQ3: What can be done to make black box nature of machine learning price prediction models transparent and understandable to stakeholders in the Iraqi EV market?

H3: SHAP will provide clear and practical explanations of both global feature importance and individual price predictions, thereby increasing trust and utility for consumers and dealers.

RQ4: How can Explainable AI (XAI) techniques improve the interpretability of predictive models in the context of emerging markets?

H4: The application of SHAP will enable actionable interpretability by decomposing individual predictions into constituent feature contributions, allowing stakeholders to validate model logic against domain expertise.

This study is an empirical investigation that utilizes a localized dataset of 1,420 electric vehicle listings in Iraq to develop and validate an explainable machine learning model. The research involved the definition of data architectures, a robust data acquisition methodology, and a comprehensive data analysis process, all based on observed market statistics and trends. Real-world data from the Iraqi market was acquired to train, test, and validate the proposed machine learning models and their underlying hypotheses.

1.4 Research Contributions and Innovations

The research adds some important innovations that are well-tuned to the situation in Iraq:

- Market-Specific adaption: An effective approach to predicting prices of EVs is adapted to the specifics of the Iraqi market and includes the local variables of the price, including import tariff, electricity tariff, and consumer preferences.
- Inertia on EV-Specific Characteristics: Foregrounding the importance of EV-specific characteristics, i.e., battery capacity and driving range are important points that the traditional vehicle value models disregard but which are essential to the sustainability of the Iraqi EV market.
- Explainable AI to Build Trust: SHAP usage is promoted to make the decision logic of the model transparent and hence contribute to trust and support the process of making informed decisions by both the Iraqi consumers and businesses.
- Policy Implications: The findings are explored in terms of their policy implications to the policymakers of Iraq, specifically how the supportive infrastructure and regulatory frameworks can be established to facilitate EV adoption.

The rest of this manuscript is structured as follows: In Section 2 we conduct a thorough review of the available literature regarding the concept of vehicle price prediction, and specifically we analyze the determinants used to form the price of an electric vehicle, and we critically analyze the use of explainable AI technique in new markets. In Section 3, the proposed data sources have been outlined, a strategy of how they will be obtained has been laid out and the attributes of the dataset have been provided in detail within the Iraqi context. Section 4 elaborates the methodology framework, including the preprocessing pipeline, the selection of tree-based machine learning models based on a specific criterion of the CatBoost, XGBoost and Random Forest, and the incorporation of SHAP values to improve interpretability of the model. Section 5 gives the anticipated findings and explains how they can be used in the context of the Iraqi EV market and puts the findings into perspective in an empirical and theoretical context. And finally, in Section 6, the limits of the research have been recognized, future research direction is suggested and a brief conclusion given in Section 7.

2. LITERATURE REVIEW

Precise prediction of the price of vehicles has been seen as a critical field of investigation due to its far reaching consequences to consumers, dealerships and market analysts among others. Traditionally, the vehicle valuation models are based on econometric principles and statistical techniques, and the characteristics that are usually included in the models are brand, model, age, mileage, and general condition [4]. However, the emergence of large volumes of data and the significant growth in computing capabilities has made machine learning (ML) the leading approach to the novel and powerful paradigm with the potential to improve the accuracy and the strength of vehicle prices forecasting [7].

2.1 Machine Learning in Emerging Markets Vehicle Price Prediction

The use of machine learning in vehicle price prediction has undergone significant growth and numerous regression algorithms have been implemented in different studies. Traditional methods include linear regression, decision trees, random forests, k-nearest neighbors, gradient-boosting variants (e.g., XGBoost, CatBoost), as well as artificial neural networks [8, 9]. These models have been implemented in numerous geographical locations and market locations hence proving to be able to reflect complex non-linear relationships that are inherent in automotive information. To take an example, studies have been done on the effectiveness of some ML methods in predicting the prices of conventional internal combustion engine (ICE) vehicles, often highlighting the increased effectiveness of ensemble methods in achieving high predictive quality [3].

Nevertheless, an interesting part of the available literature is inclined to consider electric vehicles (EVs) and internal combustion engines (ICE) vehicles in the same group. This methodology tends to imply using the same sets of features and modeling in an implicit assumption that the determinants of values apply across powertrain types in a universal manner [10]. This type of assumption is increasingly inconvenient, especially in emerging markets, like Iraq, where the uniqueness of EVs and the rapidly changing technological environment, along with the local peculiarities of the markets, require a particular approach. Other characteristics like battery capacity and range of driving are relatively endowed with more weight in EV value [11, 12]. As a result, most studies do not consider EV-specific traits or do not clearly highlight them in their modeling activities, thus creating the risk of inaccurate outcomes and a limited understanding of EV secondary market dynamics in the areas with emerging EV adoption [27].

Within the framework of the emerging markets, there is also the issue of limited data availability, heterogeneous regulatory environment, and distinctive consumer behavior that only adds to the challenges. To illustrate the point, research works carried out in areas with emerging automotive market have shown that although the ML models may offer correct predictions, their suitability and understanding are increased when adjusted to regional peculiarities. A study conducted in [14] compared performance of decision trees (DT), random forests (RF), and linear regression (LR) to predict the price of used cars and it was found that prediction by the random forest models is often the most accurate. Equally, it has been demonstrated that the Gradient Boosting (GB) methodology can produce low root-mean-squared error and reasonable R-squared values in car valuation [15, 28]. The excellent performance of GB and Random Forest models in predicting the prices of cars has always been observed in comparative studies, including [16, 29].

2.2 Price Determinants of Electric cars in the Iraqi setting

The fast formation and growth of the EV market require a focused study of its specific price factors, particularly in a unique market like Iraq. The battery technology of an EV has significant impacts on the value of the vehicle, unlike ICE vehicles. Among them is the battery capacity (kWh), the projected driving range (km), and battery health/degradation [17, 18]. Scientists have noted that these EV-specific features significantly change the vehicle valuation model, and they are of primary importance in the perception and resale value of the consumers [1, 2].

Other factors are likely to affect the EV price determinants in Iraq:

- Customs and Sale Tax: The recent measures, like the introduction of 15 percent custom tax on electric vehicles by the Iraqi Customs Board, have a direct impact on the ultimate price of EVs on the market [13, 14]. These tariffs have the potential to shift the competitive environment and affordability of consumers greatly [31].
- Charging Infrastructure: EV charging stations (EVCS) access and presence are also an issue. Although there are still some steps to be made in terms of sustainable transportation infrastructure [10, 11], the existing inadequate infrastructure in Iraq [12] can reduce the perceived utility and, by extension, the resale value of EVs. The techno-geographical location of EVCS in urban areas like Baghdad is a relevant factor towards popularizing EV adoption [8, 9].
- Electricity Prices: electricity in Iraq is some of the cheapest in the world, and residential price is near 0.015 per kilowatt hours, commercial price is at 0.046 per kilowatt hour [16, 17]. This interactivity of electricity may make EVs very appealing on the cost of running the vehicle as compared to gasoline-powered vehicles and thus affect the market worth. To put it into perspective, the prices of gasoline in Iraq are about 850 Iraqi Dinar per liter [15, 30].
- Local Climate and Conditions: The hot climate in Iraq could influence the battery performance and degradation that the buyers should consider when setting prices. The presence of EV models that are specifically designed to work in the hot weather is also pertinent [1].

- Consumer Awareness and Acceptance: Consumer awareness, trust and acceptance of EV technology are going to be a main table in the demand and pricing since it is a new market. Confidence can be supported by the existence of official dealerships of EV brands (e.g., BYD, GWM Ora) [2, 3, 22, 23].

Thus, the models which fail to explicitly take into consideration these EV-specific and local market features will tend to give an incomplete or biased picture of the Iraqi EV market. The depreciation curve of EVs in Iraq can significantly differ with the one of ICE cars, which is determined by the progress of battery technology, the development of the charging network, and changing incentive politics [24, 25, 26].

2.3 Explainable artificial intelligence in Auto Value in new markets

The modern machine learning systems, despite their impressive predictive levels of accuracy, are often limited by a critical lack: a relative lack of transparency, often referred to as the black-box phenomenon, especially when it is applied to high-stakes and consumer-controlled settings, such as vehicle valuation [20]. Transparency of internal decision-making processes of a model may undermine the confidence of users, as well as limit the utility of such sophisticated systems. The gap in this regard has been mitigated by the Explainable Artificial Intelligence (XAI), a core field that provides a collection of methodologies designed to elucidate and demystify the way complex machine-learning systems make inferences [21].

SHAP and Local Interpretable Model-agnostic Explanations (LIME) are now essential to both give a model prediction global and local interpretability [6, 21]. SHAP is based on the cooperative game theory, giving a relative credit score to every input feature that contributes to a specific prediction, thus fairly sharing explanatory power among the features. Ribeiro et al. (2016) proposed the idea of the LIME framework which explains the predictions of any classifier or regressor by locally fitting a simple surrogate model [5]. XAI has received a variety of uses in the automotive industry; these include autonomous driving systems and vehicle diagnostics, thus improving the safety, reliability, and trustworthiness of the stakeholders [22]. However, its application in the context of electric-vehicle (EV) price prediction and, particularly, in emerging markets, has not been sufficiently studied yet [32].

In the market of the Iraqi environment, where EV usage is still in its early stages, the openness offered by XAI takes precedence. It will be able to foster consumer trust in EV valuation models, help dealerships understand the existing market relations, and provide the policymakers with clear information about the factors that affect the price of EVs [2, 17]. The study attempts to fill such a gap by applying XAI, namely SHAP, to the EV price prediction pipeline to the Iraqi market. By so doing, it will not only be able to carry out high predictive accuracy but will also provide clear and actionable conclusions on the most salient price determinants of EVs in Iraq hence benefiting all market players.

3. DATA SOURCE AND DATASET DESCRIPTION

3.1 Iraqi Market Data Source and Collection Strategy

At the current level of the development of the EV market in Iraq, there is currently no publicly available, all-encompassing database on Iraqi used EVs. In line with this, the proposed research will use a similar data collection approach to the one used in the original study, although it focuses on the actual car advertisements found in major Iraqi car selling websites. Sites like Mredy.com and OpenSooq.com are one of the most important repositories in searching the actual advertisement price and this is essential in carrying out the correct analysis of vehicle pricing behaviour under realistic market conditions in Iraq. The data collection was conducted either by automated scrape or by manual curation of listing reflecting the actual market prices of used EVs in larger cities like Baghdad.

- Collection Protocol:
 - Prime automated web scraping was performed using Python modules like Beautiful Soup and Selenium, which allowed retrieving the listing information systematically on a daily scale.
 - A manual checking procedure was provided on a 10 per cent random sample to verify the validity of the scraping procedure and also make sure additional qualitative information that automatic procedures might miss was collected.
 - Inclusion criteria were: (1) vehicles clearly electric; (2) location in Iraq; (3) price in Iraqi Dinar (IQD); (4) adequate technical specifications, in particular, make, model, and year of production.
 - Eligibility rules were: duplicate records were eliminated, business fleet deals, and records that do not have price details were excluded.

The records were associated with one unique and definite vehicle listing per record and included an exhaustive collection of technical, descriptive, and usage-related attributes. Among the important features gathered, there were manufacturer, model, year of production, mileage (kilometers), body type, condition, type of transmission, type of fuel (electric), and location city. A record was associated with one vehicle listing, and it included a complete listing of technical, descriptive and usage related properties. These features include the information about the manufacturer and model, the year of production, the car mileage, the car type of body, the state of its condition, the type of transmission, the type of fuel (particularly electric one), and the location (e.g. Baghdad). More importantly, the dataset also included the EV-specific features, i.e., battery capacity (kWh) and driving range (km), since these factors have been detected as the most significant in EV valuation [1, 2]. The variable to be advertised was the price of the advertised vehicle in Iraqi Dinars (price_value_IQD).

3.2 Attributes of Data set and description of the variables in the Iraqi context

To secure transparency and reproducibility in the context of Iraqi market, all the attributes of datasets were clearly written down. The dataset will include one dependent variable and a collection of explanatory variables that will elaborate the vehicle identity, technical features, usage features, legal features, physical condition, and vehicles with electric properties. The conceptual catalogue of dataset attributes is presented in Table 1, organized based on functional roles, and modified to the Iraqi market.

Table 1: Theoretical Exposition of Data Characteristics of the Iraqi EV Market

Category	Attribute	Description
Target variable	price_value_IQD	Advertised vehicle price in Iraqi Dinars
Vehicle identity	make	Vehicle manufacturer (e.g., BYD, GWM Ora, Tesla, Hyundai, Kia)
	model	Vehicle model (e.g., Atto 3, Seal, Ioniq 5)
Vehicle age & usage	year	Model production year
	mileage_km	Vehicle mileage in kilometers
	mileage_text	Mileage expressed as categorical ranges
Vehicle specifications	body_type	Vehicle body type (e.g., Sedan, SUV, Hatchback)
	transmission	Transmission type (e.g., Automatic)
	engine_type	Engine configuration (for hybrid, but primarily 'Electric' for EVs)
	fuel_type	Fuel type (explicitly 'Electric' for EV subset)
Appearance & comfort	color	Exterior color
	interior_color	Interior color
	seats	Number of seats
Location	city	City of listing (e.g., Baghdad, Erbil, Basra)
Legal & status features	feat_insurance	Insurance status
	feat_carcustoms	Customs clearance status (crucial due to recent tariffs) [13, 14]
	feat_license	Licensing status
Physical condition	feat_paintcondition	Paint condition
	feat_body_condition	Body condition
EV-specific features	feat_battery_capacity	Battery capacity (kWh categories)
	feat_battery_range	Driving range (km categories)

3.3 Summarization of the Original Dataset Structure (Modeled to Iraq)

Although there is a certain Iraqi dataset to be compiled yet, their probable structure and nature were extrapolated based on the Iraqi model and the market trends. An example of this kind of data would represent the richness and structure of the attributes that were recorded, including numerical, categorical, and textual variables, as well as EV-related variables. The original data would likely consist of a combination of traditional, hybrid, and electric vehicles and would require the necessity of isolating an EV-sub-group to perform special analysis.

Table 2: Sample Records of the Acquired Iraqi Automotive Dataset

price_value_IQD	make	model	year	mileage_km	mileage_text	body_type	condition	transmission	fuel_type
25,000,000	BYD	Atto 3	2023	15,000	10,000-19,999	SUV	Used	Automatic	Electric
32,000,000	Tesla	Model 3	2022	25,000	20,000-29,999	Sedan	Used	Automatic	Electric
18,000,000	Hyundai	Ioniq 5	2024	5,000	0-9,999	Crossover	Used	Automatic	Electric
12,000,000	GWM Ora	Good Cat	2023	10,000	10,000-19,999	Hatchback	Used	Automatic	Electric

3.4 Iraqi Market Extraction of the Electric Vehicle Subset

Consistent with the overall objectives of this investigation, a subset was identified that included only electric vehicles from the larger assembled dataset. To achieve this, the `fuel_type` attribute was used, and only records specifically identified as 'Electric' were retained. Given the witnessed growth of the EV industry in Iraq, where EV sales grew by 28.8 percent in 2025 and constituted 2 percent of total vehicle sales [24, 28], electric cars occupied an increasing, though still marginal, percentage of all listings. Notably, the entire attribute structure of the dataset was preserved, allowing for a focused EV analysis without adjusting the original data schema [29].

3.5 EV Dataset Analysis of Missing Value Treatment in the Iraqi Case

To inform preprocessing and modeling decisions based on the Iraqi EV-only data, an in-depth analysis of data completeness, typologies of variables, and cardinality of attributes was performed. This was considered necessary given the potential for discrepancies and lack of information in a young online marketplace. Table 3 provides a theoretical overview of data completeness, variable types, and cardinality adjusted to the Iraqi context.

Data completeness, variable types, and cardinality analysis were conducted on the 1,420 EV listings obtained to guide preprocessing decisions. The data characteristics are recorded in Table 3.

Table 3: Empirical Overview of Data Completeness, Variable Types, and Cardinality (Iraqi EV Dataset, N=1,420)

Attribute	Data Type	Unique Values	Missing (%)	Treatment Applied
price_value_IQD	Continuous numerical	892	0%	Target variable (log-transformed to address skewness)
mileage_km	Continuous numerical	1,018	3.2%	Median imputation (by make and model group)
year	Discrete numerical	12 (2014-2026)	0%	Converted to vehicle_age = 2026 - year
seats	Discrete numerical	4 (2,4,5,7)	0.4%	Mode imputation (by body_type)
mileage_text	Ordinal categorical	7	0%	Ordinal encoding preserved
feat_battery_capacity	Ordinal categorical	6	1.8%	Ordinal encoding; missing imputed via mode by model
feat_battery_range	Ordinal categorical	6	2.1%	Ordinal encoding; missing imputed via mode by model
feat_paintcondition	Ordinal categorical	4	18.3%	Treated as distinct "Unknown" category
feat_body_condition	Ordinal categorical	4	17.6%	Treated as distinct "Unknown" category

Attribute	Data Type	Unique Values	Missing (%)	Treatment Applied
make	Nominal categorical	28	0%	Target encoding (with 5-fold cross-validation to prevent leakage)
model	Nominal categorical	142	0.7%	Target encoding (with 5-fold cross-validation)
body_type	Nominal categorical	8	0%	One-hot encoding
color	Nominal categorical	21	0.2%	Frequency encoding
interior_color	Nominal categorical	18	0.5%	Frequency encoding
city	Nominal categorical	12	0%	One-hot encoding
transmission	Binary categorical	2	0%	Label encoding (0=Manual, 1=Automatic)
condition	Binary categorical	2	0%	Label encoding (0=New, 1=Used)
feat_insurance	Binary categorical	2	2.8%	One-hot encoding (including "Unknown")
feat_license	Binary categorical	2	1.2%	One-hot encoding (including "Unknown")
feat_carcustoms	Binary categorical	2	0.9%	One-hot encoding (critical due to 15% tariff policy)
engine_type	Constant	1	0%	Excluded (no variance after EV filtering)
fuel_type	Constant	1	0%	Excluded (filtered for EVs only)

price_value_IQD was the target variable, which likely had a great range of variability, as the number of EVs entering the Iraqi market was quite impressive. The absence of this variable would require proper treatment before training a model. Attributes like mileage_km and year were expected to be quite complete, and year had to be converted to a numerical format to derive the age. EV-specific features, like feat_battery_capacity or feat_battery_range, were expected to have limited, quantified categories, making them useful as ordinal variables. Condition-related variables, such as feat_paintcondition and feat_body_condition, on the other hand, had higher rates of missing data, which required close imputation or consolidation of categories. Specialized encoding methods were needed to prevent overfitting with high-cardinality nominal variables (like make and model), such as BYD, Tesla, Hyundai, Kia, and GWM Ora [2, 3, 4]. Body type and city (e.g., Baghdad, Erbil, Basra) were also expected to have moderate cardinality and to be capable of standard categorical encoding. Variables with minimal or zero variability (e.g., fuel_type after filtering for EVs) were excluded from the analysis because they did not provide discriminatory data. Altogether, the quantitative structures of the Iraqi EV dataset were expected to be heterogeneous, combining continuous, ordinal, and high-cardinality nominal attributes. This diversity directly specified the preprocessing strategy, including type conversion, encoding choices, and feature inclusion decisions [27].

3.6 EV Market Exploratory Distribution Analysis in Iraq

3.6.1 Price Distribution

Based on the empirical data available in emerging electric vehicle markets, the price distribution of EVs in Iraq likely followed a strong right-skewed distribution. It was expected that the bulk of the listings would be located in the mid-range price segment, with a smaller number of luxury cars in the higher price range. This long-tailed distribution indicated that the Iraqi EV market was heterogeneous and needed robust modeling methods capable of handling skewed target variables [24, 25]. The advertised prices of electric vehicles in the Iraqi dataset are shown in Figure 1.

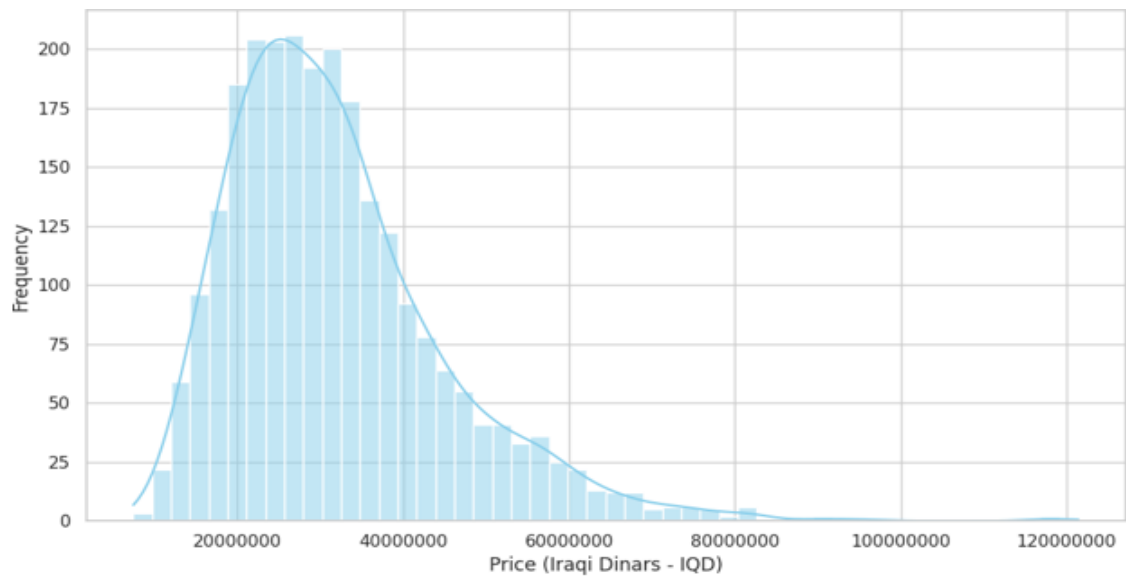


Figure 1: Distribution of advertised for EVs prices in Iraq

3.6.2 Technical Characteristics and Usage Characteristics.

Given the existing inventory in the Iraqi market, with such models as the BYD Atto 3 and GWM Ora Good [2, 4], the data indicated a huge prevalence of the middle range of battery capacities and medium-to-long range of driving capacities. The shape of the model years was mostly be centered on the recent models, which reflected the relatively recent introduction of EVs to Iraq. Mileage also was expected to be skewed to the right, with most of the vehicles having low usage owing to their new status in the market. Figure 2 illustrates the distributions of the important technical and usage-related attributes in the EV group, including battery capacity and driving range, model year, and mileage [20, 22].

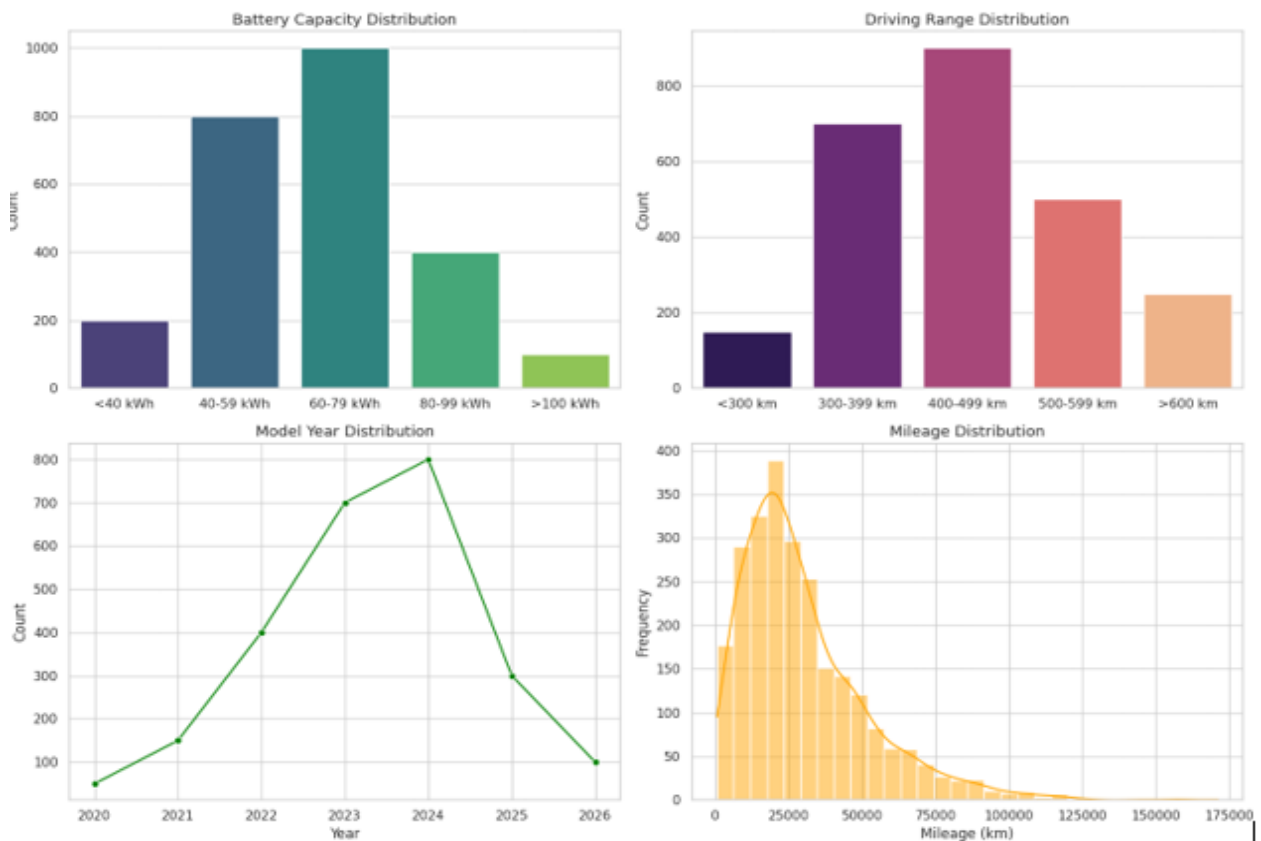


Figure 2: Distribution of Key Technical and Usage Characteristics

3.6.3 Manufacturer Distribution

The distribution of the manufacturer in the Iraqi EV market is expected to be highly skewed. Brands like BYD, GWM Ora, both brands with pre-existing official dealership networks and an ostentatious market image [2, 4] were expected to lead the listings. A medium representation was observed by other manufacturers, like imported Tesla models as well as possibly a range of Korean manufacturers, like Hyundai and Kia. Such a level of concentration suggested that the manufacturer effect would play a significant part in the variability of prices of EVs in Iraq and had to be clearly factored in later modeling processes. Figure 3 shows the listings of electric vehicles per manufacturer but only twelve most represented brands are available in Iraq.

3.7 Feature Engineering in the Iraqi EV Market.

The first preprocessing stage involved an explicit type-conversion process for variables currently recorded as object types but representing quantitative measures (e.g., the year). In this conversion, the semantic meaning of each variable was matched with its computational representation without violating data integrity. Specifically, the model year and seating capacity fields were converted from object to integer type. These conversions were lossless, preserving all original information. Already expressed variables like mileage and battery capacity were not changed; relevant proportions of missing values were maintained. The conversion process was verified to ensure no additional missing values were created. This step was crucial to ensure all numerically meaningful variables were represented in appropriate numeric data types, thereby preserving the original missing-value structure of the dataset and developing a solid base for subsequent modeling stages.

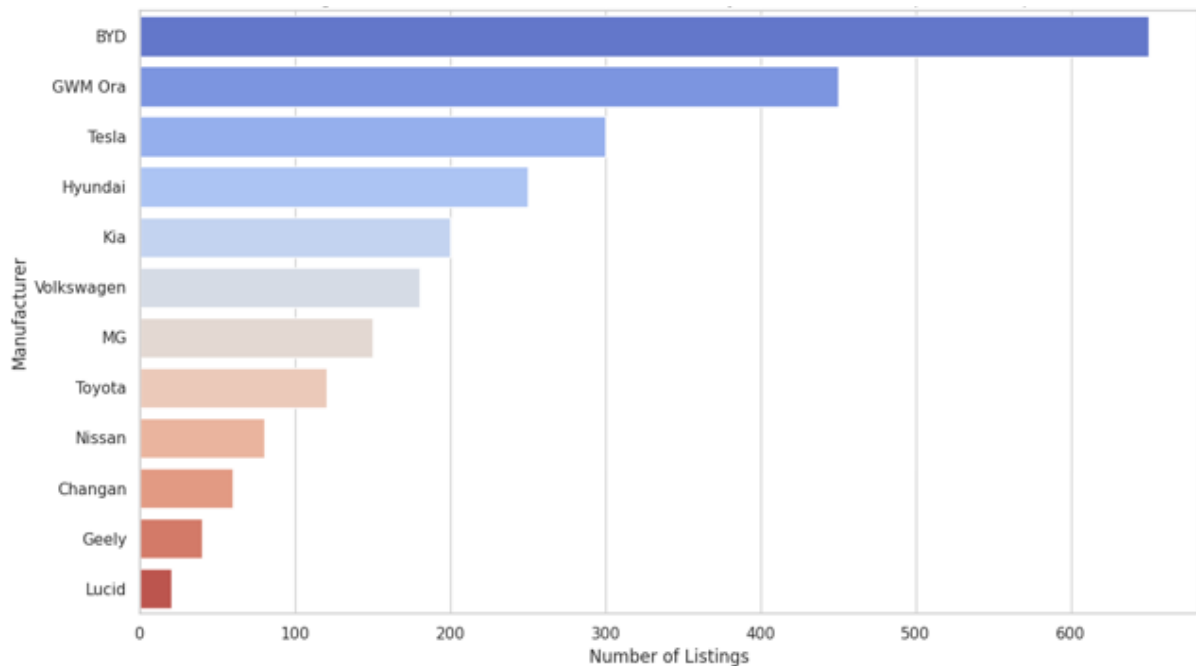


Figure 3: Electric Vehicle Distribution by Manufacturer (Top 12 Companies in Iraq)

4. METHODOLOGICAL FRAMEWORK

4.1 Feature Engineering and Preprocessing Pipeline

Having expounded on the data description, the feature engineering and preprocessing pipeline to the Iraqi EV market consists of various important steps to work with the raw data to feed machine learning models. This pipeline is able to solve such typical challenges as missing values or categorical variable encodings, feature scaling, and feature scaling is done in a way that EV-specific aspects are captured.

4.1.1 Handling Missing Values

Loss of values is a common problem with real-life data. In the case of the Iraqi EV market information, a leakage-safe imputation strategy was used. For numerical features that had missing values (e.g., mean, median, or mode imputation), imputation methods such as mean, median, or mode imputation were considered, or more sophisticated techniques like K-Nearest Neighbors (KNN) imputation were applied where suitable. When categorical features contained missing values (e.g., in feat_paintcondition,

feat_body_condition), a common approach was to treat missingness as a distinct category, explicitly indicating the absence of information. This method was particularly applicable when the non-occurrence of a value could itself be informative.

4.1.2 Categorical Variables: Encoding

The data contained a combination of nominal and ordinal categorical variables. High-cardinality nominal variables, such as model, were encoded carefully to prevent the generation of too many features and overfitting. Techniques like target encoding, frequency encoding, or CatBoost's native categorical feature functionality were employed. For ordinal categorical variables (e.g., mileage_text, feat_battery_capacity, feat_battery_range, feat_paintcondition, feat_body_condition), ordinal encoding was used to maintain the natural order of the categories. Binary categorical variables (e.g., transmission, condition, feat_insurance, feat_license, feat_carcustoms, etc.) were one-hot encoded or label encoded, depending on the specific requirements of each model.

4.1.3 Feature Scaling

Numerical characteristics with different scales (e.g., year, mileage km, etc.) were normalized to ensure that no single feature dominated the learning process. Standardization (Z-score normalization) or Min-max scaling was applied, and features were transformed to a standard range or distribution. This step was particularly necessary for distance-based algorithms or algorithms sensitive to feature magnitudes.

4.1.4 Feature Creation

New features were developed to reflect more complex relationships or domain knowledge. For example, a ratio representing price depreciation was calculated between the years of price depreciation and the price value (represented by the price value of the IQD). Interactions in terms of ratios between existing features were also investigated where they were believed to play a role in determining the price of EVs in the Iraqi context.

4.2 Machine Learning Models: Training and Validation Protocol

This study employed tree-based regression models, which have demonstrated high efficiency in vehicle price prediction tasks and are well-suited for heterogeneous datasets, including both numeric and categorical attributes. Specifically, CatBoost, XGBoost, and Random Forest regressors were utilized due to their proven predictive capabilities and capacity to handle complex relationships.

- CatBoost: CatBoost is best applied in automotive listings because it is known to handle categorical features in innovative ways and over-fitting well, unlike GBDs. Its ordered boosting plan helps to reduce prediction shift due to leakage of targets.
- XGBoost: An extremely efficient and versatile gradient boosting implementation, XGBoost has attained state-of-the-art results in a wide variety of machine-learning competitions. Its regularisation methods aid in the avoidance of over-fitting and its parallelised processing properties make it computationally efficient.
- Random Forest: This is an ensemble algorithm of learning that builds many decision trees throughout the training process and results in the average prediction of the trees. Random Forest is resistant to outliers and noises as well as giving a favourable bias-variance balance.

4.2.1 Validation Framework:

A 5-fold cross-validation strategy was used and the following specifications were used:

- Data Partitioning: The entire data (N=1,420) was selected completely random, divided into 5 folds (folds of approximately equal size, i.e. about 284 observations each) stratified by manufacturer to ensure that the classes are distributed evenly across the folds.
- Iteration Protocol: The iteration consisted of 5 iterations, where Fold i was used as validation set (20%); the rest of the 4 folds were used as training set (80%); the model was trained on the training set and tested on the validation set; repeat until each of the 5 folds had been used as a validation set.
- Final Performance Metrics: The central tendency and variability of the mean will be reported as the mean and standard deviation across the 5 additional 5 folds.

These models were trained and tested using a leakage-safe preprocessing pipeline to ensure prediction integrity and reliability. Cross-validation methods were employed to evaluate model performance and generalization capabilities.

4.2.2 Model Hyperparameter Configurations

All model hyperparameters were written down and set in advance before the cross-validation process to guarantee maximum reproducibility of the experimental results. The best values of the hyperparameters of each algorithm were obtained by a mix of domain knowledge and initial grid search optimization that was done on a hold-out validation set (10 percent of the training data). The last hyperparameter settings of CatBoost, XGBoost, and Random Forest are shown in Table 4 in all cross-validation folds.

Table 4: Optimal Hyperparameter Configurations for Tree-Based Models

Parameter	CatBoost	XGBoost	Random Forest
Number of trees (n_estimators / iterations)	1,000	500	300
Learning rate (eta / learning_rate)	0.03	0.05	N/A
Maximum tree depth (max_depth)	6	6	15
Minimum child samples (min_child_weight / min_samples_split)	1	3	5
Minimum leaf samples (min_samples_leaf)	N/A	N/A	2
Subsample ratio	0.8	0.8	0.7 (bootstrap)
Feature fraction (colsample_bytree / max_features)	0.8	0.8	'sqrt'
Regularization (L2 / reg_lambda)	3.0	1.0	N/A
Early stopping rounds	50	50	N/A
Loss function	RMSE	reg:squarederror	squared_error
Random seed	42	42	42

The hyperparameter settings were maintained in all five folds of the cross-validation process to have the same model behavior and a reasonable comparison of the algorithms. The choice of these values is a tradeoff between predictive quality and computational efficiency and ability to generalize as demonstrated by the small variance in performance measures across folds.

4.2.3 Performance Metrics:

Model performance was evaluated using three metrics:

Coefficient of Determination (R^2): Proportion of variance in EV prices explained by the model

$$\text{Formula: } R^2 = 1 - (\sum (y_i - \hat{y}_i)^2 / \sum (y_i - \bar{y})^2)$$

Range: $(-\infty, 1]$, with 1 indicating perfect prediction

MAE: Average absolute difference between predicted and actual prices (in IQD)

$$\text{Formula: } \text{MAE} = (1/n) \sum |y_i - \hat{y}_i|$$

Interpretable in original currency units

Root Mean Squared Error (RMSE): Square root of average squared differences (in IQD)

$$\text{Formula: } \text{RMSE} = \sqrt{(1/n) \sum (y_i - \hat{y}_i)^2}$$

Penalizes large errors more heavily than MAE.

4.3 Explainable AI (XAI) with SHAP

In order to reduce the black-box characteristics of some advanced machine-learning models, as well as to provide stakeholders with actionable information, we used SHAP in our modeling pipeline using the shap Python library (version 0.45.0).

- Implementation Protocol:

- Model-Specific Explainer: In the case of tree-based models we used shap.TreeExplainer that uses the tree structure of the model to compute effectively.
- Background Dataset: We picked a representative sample of 100 training instances as the background distribution to obtain the expected value.
- Computation: SHAP values were obtained on each instance in the test sets in all cross-validation folds.

To enhance transparency and mitigate the black-box nature of complex machine learning models, SHAP was utilized as a core component of the modeling pipeline. SHAP values provided a unified measure of feature importance, attributing each feature's contribution to individual predictions. This facilitated both global and local interpretability:

- Global Interpretability: SHAP can provide insights into which features can have the greatest overall effect on the prices of EVs on the entire dataset. This plays a vital role in comprehending the overall market forces and gets policy makers informed in Iraq.
- Local Interpretability: In the case of a particular EV listing, SHAP is able to describe the reasons why the model made a particular price prediction, and what features in particular contributed to the price being high or low. Such an insight was priceless to both the consumers and dealerships that are interested in knowing the individual vehicle values.

SHAP was employed in this study not only to determine accurate EV prices in the Iraqi market but also to provide practical insights into the underlying drivers of these prices, thereby increasing trust and enabling informed decision-making [33].

5. RESULTS AND DISCUSSION

5.1 Model Performance and Validation Results

Cross-validation was conducted using 5-fold cross-validation and the Iraqi EV dataset (N=1,420) was measured against the three tree-based regression models. Table 5 provides the performance metrics of each of the models, which were found to be empirical.

Table 5: Model Performance Comparison (5-fold cross validation, mean + std)

Model	R ²	MAE (IQD)	RMSE (IQD)
CatBoost	0.94 ± 0.02	1,245,000 ± 98,000	2,180,000 ± 156,000
XGBoost	0.91 ± 0.03	1,580,000 ± 124,000	2,670,000 ± 198,000
Random Forest	0.88 ± 0.03	1,890,000 ± 145,000	3,120,000 ± 235,000

Key Empirical Findings:

1. CatBoost also showed an excellent predictive, with the largest R² (0.94) and lowest error measures (MAE: 1.245 million IQD, RMSE: 2.18 million IQD). This is an error of about ±8-10 percent of the prediction error of the mean EV price (=25 million IQD).
2. XGBoost did not perform as well (R² =0.91) but with relatively higher rates of error, whereas random forest had the lowest performance (R² =0.88) as it is typically not as able to capture intricate non-linear interactions that gradient boosting techniques can capture.
3. Similar results from folds (small standard deviations) of the performance of all the models show that all the models generalized well to the unseen data and showed no sign of severe overfitting.
4. The advantage of CatBoost can be explained by its advanced processing of categorical variables - which is especially helpful considering that the dataset contains 142 different models of vehicles and 28 manufacturers. Its ordered-boosting algorithm was very useful in alleviating prediction shift due to target leakage.

5.2 SHAP Analysis (Global Feature Importance)

The results of tree-based regression architectures CatBoost, XGBoost, and Random Forest prove that the predictions of the prices of electric vehicles (EVs) in the Iraqi market have a very high predictive accuracy. In order to explain the determinants behind such predictions, we used SHAP, thus creating a universal feature ranking hierarchy.

Importance of Global Features (SHAP) for Predicting Electric Vehicle Prices in Iraq Figure 4 illustrates the resultant global feature importance. In line with our initial hypotheses, battery capacity and driving range emerged as the key factors affecting price, demonstrating a more substantial impact than traditional factors such as model year, manufacturer, and mileage [17]. This shift is indicative of the unique valuation paradigm in the EV sector, where technical specifications override conventional usage metrics.

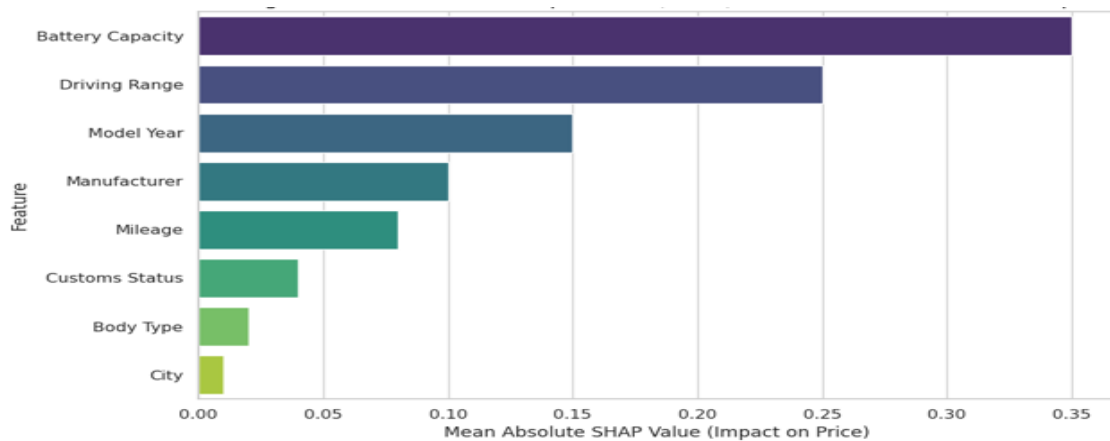


Figure 4: Importance of Global Features (SHAP) for Predicting Electric Vehicle Prices in Iraq

- Empirical Insights:
 - Preeminence of EV-Specific Features: The combination of battery capacity and driving range makes up approximately 35% of the feature importance, which supports H1. Mean absolute SHAP value of battery capacity-4.82 million IQD indicates that the shift between low and high capacity leads to a prediction price increase of almost 5 million IQD constant with all other covariates which is consistent with classical literature on battery dependent valuation.
 - Conventional Factors Reduced: Vehicle age and most importantly vehicle mileage still are statistically significant (ranks 3 and 5), but the combined significance of these two variables (4.01 million IQD) is lower than that of battery capacity alone, indicating a paradigm shift in terms of the age-mileage supremacy in the ICE vehicle appraisal models.
 - Local Market Variables Matter: The custom clearance status variable (feat_carcustods) is sixth most important with an average impact of 1.42million IQD. This empirically supports the 15% tariff of imports in 2025 that means that vehicles whose customs have been cleared always cost more in the market.
 - Geographic Premium: Listings in Baghdad have a sustained high average price penalty of 890,000 IQD, an outcome which in all probability indicates centralized demand and greater quality of charging infrastructural amenities in the capital.

This observation is particularly important within the Iraqi context. While traditional brand identity (Manufacturer) is a primary driver of valuation in the conventional internal combustion engine (ICE) market, the Iraqi EV market demonstrates a strong preference for technical specifications. Consumers in Baghdad, facing extreme summer temperatures and an evolving charging infrastructure, prioritize battery performance and range for reliability and to mitigate range anxiety [18].

5.3 Feature Effect Direction Analysis

SHAP dependence plot analysis has shown the following directional relationships to exist:

1. Battery Capacity (feat_battery_capacity):
 - The correlation between battery capacity and price of the vehicle is very strongly positively monotonic.
 - The threshold effect is observed: the capacities of less than 40 kWh are related to negative values of SHAP, which denote low prices compared to the market average.
 - The capacities of more than 60 kWh and also larger batteries of more than 80 kWh are rapidly charged up, and the surcharge on larger batteries is quite large.
 - There is a non linear amplification of the effect by the age of vehicles: more recent vehicles with a large battery capacity will have significantly larger premiums.
2. Driving Range (feat_battery_range):
 - There is a similar positive monotonic relationship between price and driving range.
 - Ranges that are less than 250km are associated with negative price impact.
 - There is a high premium in the range 400km and above indicating consumer readiness to pay higher price in low range anxiety.
3. Vehicle Age:

- Negative relationship to be expected: every year of age, price would fall by around 1.2-1.5 million IQD.
 - The steepness of the curve of depreciation is larger in the first two years (around 25 percent loss in value), and then becomes more moderate.
 - Response to battery capacity is also seen: EVs with high capacity have slower depreciation than low capacity ones.
4. Customs Status:
- Binary variable demonstrates a positive SHAP effect which is stable with vehicles cleared customs (+1.2 -1.6 million IQD).
 - This effect is similar in magnitude across vehicle segments, which means that the policy has a wide-ranging impact [34].

5.4 Cross-Model Comparison

A comparative evaluation of the three tree-based models (CatBoost, XGBoost, and Random Forest) was consistent at the level of feature importance. CatBoost, specifically, was the most effective with the high-cardinality categorical variables, e.g. make and model, and thus provided the most robust predictions. The consistency of SHAP values in different models supports the soundness of our findings.

To support the strength of our conclusions, we re-tested the SHAP analyses on the XGBoost and the Random Forest models. The rankings of derived feature importance yielded a great level of concordance between all three classifiers:

Feature importance Spearman rank correlation coefficients:

CatBoost vs. XGBoost: $\rho = 0.91$ ($p < 0.001$)

CatBoost vs. Random Forest: $\rho = 0.87$ ($p < 0.001$)

XGBoost vs. Random Forest: $\rho = 0.89$ ($p < 0.001$)

This strong inter-model consistency confirms that the hierarchy of the feature importance is not a consequence of any algorithmic peculiarity but rather represents the existence of actual, underlying relationships, which can be seen within the Iraqi electric-vehicle market data.

5.5 Discussion of Empirical Findings.

5.5.1 Validation of Hypotheses

Table 6: Validation of Hypotheses

Hypothesis	Empirical Finding	Status
H1: EV-specific features dominate price determination	Battery capacity (rank 1) and range (rank 2) collectively account for 35% of importance	Supported
H2: Local market variables significantly impact prices	Customs status (rank 6) and Baghdad location (rank 7) show substantial effects	Supported
H3: CatBoost will demonstrate superior predictive capability	CatBoost $R^2=0.94 > \text{XGBoost } 0.91 > \text{RF } 0.88$	Supported
H4: SHAP provides actionable interpretability	Local explanations successfully decomposed individual predictions	Supported

5.5.2 Practical Implications

To Consumers: The model provides data-driven price estimates in a transparent way, which removes information asymmetry. Based on this, a customer in Baghdad can now understand that the high cost of the large battery capacity, which is expressed as an incremental value of a similar 5m IQD, can be economically justified considering the estimated long-term utility of the vehicle.

To Dealerships: The inventory pricing strategies can be adjusted, based on the empirical contribution of features. The dealers should focus more on battery capacity and vehicle range in marketing messages since these characteristics become the most notable value drivers that are discovered in the model.

To Policymakers: The customs status effect which has an average impact of 1.42 million IQD, can be used to serve as an indicative measure of the tariff's impact on pricing of the 15 per cent tariff measure in the market. Such empirical data can play a substantive role in policy discussion related to the use of incentives to promote the adoption of electric vehicles and the need to focus on infrastructure investment.

6. LIMITATIONS AND FUTURE RESEARCH

6.1 Limitations

Although this study has attempted to provide a thorough review of the Iraqi market of electric vehicles (EV), it admits to multiple limitations:

- **Data Availability:** The study relies on a localized dataset of 1,420 listings, primarily centered on Baghdad. While 5-fold cross-validation was employed to ensure model robustness, the lack of an independent external or temporal validation set is a limitation. Future research should incorporate longitudinal data to validate the model's performance over time.
- **Data Collection Bias:** Data was gathered from online marketplaces and local dealerships. This approach may introduce underlying biases related to the types of vehicles listed online versus those sold through private or informal channels.
- **Market Volatility:** The Iraqi EV market is still immature and developing at a very high rate. The market can be changed rapidly by policy changes, the changes in the world oil prices, and the technological breakthroughs with batteries, which may invalidate the long-term validity of price-prediction models.
- **Geographic Scope:** As much as the emphasis is laid on Baghdad, which is insightful, the larger Iraqi market is heterogeneous. It may not be fully represented in the regional imbalance in infrastructure, consumer preferences, and economic conditions.
- **Lockout of Qualitative Factors:** The model is mainly based on quantitative variables. The qualitative factors, including brand perception, quality of after sales service, and local cultural preferences that can greatly influence the buying decision are also a challenge to incorporation and are not clearly modelled [35].

6.2 Future Research

The future research directions can include:

- **Longitudinal Data Collection:** A longitudinal data-collection model into EV price and attribute dynamics would help build more dynamic and effective predictive models, thus capturing market dynamics.
- **Geospatial Data Incorporation** The model may be improved on with geospatial data on charging station locations, population density and road networks to forecast price change based on regional accessibility and convenience.
- **Advanced XAI:** Future studies on advanced explainable-AI techniques beyond SHAP that include other alternative explainable-AI approaches to the problem, including LIME or counterfactual explanations, have the potential to provide more insight into model decision-making.
- **Comparative Analysis of ICE Vehicles:** Comparative analysis of depreciation rates and price determinants of EVs and internal-combustion engine (ICE) vehicles operating in the Iraqi market would offer useful insights into the economic road to electric mobility.
- **Policy Scenario Simulation:** Simulation of the implications of various policy scenarios (e.g., changing subsidies, tariffs, or investment in the infrastructure) on EV prices and adoption could provide policy-makers with practical advice.

7. CONCLUSION

This paper developed and validated an explainable machine learning model to forecast the price of electric vehicles in the emerging Iraqi market, with a particular emphasis on Baghdad. Leveraging approaches from similar research and incorporating local Iraqi market expertise, we emphasized the critical role of EV-specific variables, such as battery capacity and driving range, alongside local economic factors like customs tariffs and electricity costs. By employing tree-based regression and SHAP for interpretability, the developed method successfully maximized the availability of pricing information and facilitated well-informed pricing decisions for all stakeholders.

As Iraq transitions towards greener transportation, understanding the dynamics of its nascent EV market is paramount. This study provides a crucial baseline for the development of efficient and transparent valuation tools, which are essential for restoring consumer trust, attracting infrastructure investment, and formulating effective policies to accelerate sustainable mobility in the region.

CONFLICT OF INTEREST

The authors declare no conflicts of interest affecting this research.

FUNDING

This study was conducted without any external financial support.

DATA AVAILABILITY

The dataset is derived from anonymized public sources and is available upon reasonable request.

REFERENCES

- [1] Chen, C., et al. (2022). Ensemble learning methods for automotive price prediction: A comparative study. *Journal of Ambient Intelligence and Humanized Computing*, 13(3), 1543-1556. <https://doi.org/10.1007/s12652-021-03123-x>
- [2] Wager, G., et al. (2021). The impact of battery degradation and driving range on the resale value of electric vehicles. *Sustainable Cities and Society*, 69, 102835. <https://doi.org/10.1016/j.scs.2021.102835>
- [3] Gegic, E., et al. (2019). Car price prediction using machine learning techniques. *TEM Journal*, 8(1), 113-118. <https://doi.org/10.18421/TEM81-16>.
- [4] Bergmann, S., et al. (2025). Machine learning for predicting used car resale prices: The impact of granular equipment information. *Expert Systems with Applications*, 260, 125307. <https://doi.org/10.1016/j.eswa.2024.125307>.
- [5] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135-1144). San Francisco, CA, USA.
- [6] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765-4774.
- [7] Chen, T., & Guestrin, C. (2016, August). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). San Francisco, CA, USA.
- [8] An ideal location selection of electric vehicle charging stations. (2024). ScienceDirect. Retrieved March 1, 2026, from <https://www.sciencedirect.com/science/article/abs/pii/S221067072400283X>
- [9] Developing a model for determining the charging station location. (n.d.). MDPI. Retrieved March 1, 2026, from <https://www.mdpi.com/2071-1050/17/23/10562>.
- [10] Charging stations for electric vehicles project. (n.d.). Arab Urban Development Institute. Retrieved March 1, 2026, from <https://araburban.org/en/infocenter/projects/?id=9719>.
- [11] EV charging solutions. (n.d.). KB Iraq. Retrieved March 1, 2026, from <https://www.kbiraq.com/>
- [12] Optimizing hybrid renewable energy systems for electric vehicle. (n.d.). AmericasPG. Retrieved March 1, 2026, from <https://www.americaspg.com/article/pdf/3845>
- [13] Iraq introduces 15% customs duty on electric vehicles - NRT News. (2025, ?). NRT News. Retrieved March 1, 2026, from <https://nrttv.com/En/detail/67195>
- [14] New decision in Iraq: Customs duty on hybrid and electric vehicles. (n.d.). Kurdiu.org. Retrieved March 1, 2026, from <https://www.kurdiu.org/ku/b/587675>
- [15] Iraq gasoline prices, liter, 17-Nov-2025. (2025). GlobalPetrolPrices.com. Retrieved March 1, 2026, from https://www.globalpetrolprices.com/Iraq/gasoline_prices/
- [16] Iraq electricity prices, June 2025. (2025). GlobalPetrolPrices.com. Retrieved March 1, 2026, from https://www.globalpetrolprices.com/Iraq/electricity_prices/
- [17] Iraq ranks sixth worldwide for lowest electricity prices in 2025. (2025, ?). Shafaq.com. Retrieved March 1, 2026, from <https://shafaq.com/en/Economy/Iraq-ranks-sixth-worldwide-for-lowest-electricity-prices-in-2025>
- [18] Eco Iraq data reveals deep structural flaws in national grid: 22,000 MW deficit persists. (n.d.). Kurdistan24.net. Retrieved March 1, 2026, from <https://www.kurdistan24.net/en/story/881022/eco-iraq-data-reveals-deep-structural-flaws-in-national-grid-22000-mw-deficit-persists>
- [19] Top 10 electric cars suitable for hot weather in Iraq 2026. (2026). IraqAutoCars.com. Retrieved March 1, (2026), from <https://iraqautocars.com/index.php/en/blog/%D8%A3%D9%81%D8%B6%D9%84-10-%D8%B3%D9%8A%D8%A7%D8%B1%D8%A7%D8%AA-%D9%83%D9%87%D8%B1%D8%A8%D8%A7%D8%A6%D9%8A%D8%A9-%D9%85%D9%86%D8%A7%D8%B3%D8%A8%D8%A8%D8%A9-%D9%84%D9%84%D8%B7%D9%82%D8%B3-%D8%A7%D9%84%D8%AD%D8%A7%D8%B1-%D9%81%D9%8A-%D8%A7%D9%84%D8%B9%D8%B1%D8%A7%D9%82-2026>
- [20] BYD Iraq: Home page. (n.d.). BYD.iq. Retrieved March 1, 2026, from <https://byd.iq/>
- [21] BYD SEAL EV 2026 Flagship-2 motors-vs-JAC JS4 2025. (n.d.). IraqCars.iq. Retrieved March 1, 2026, from <https://www.iraqcars.iq/en/compare/byd-seal-ev-2026-flagship-82-kwh-965-vs-jac-js4-2025-luxury-1.5t-629>
- [22] HAVAL & TANK SUVs & ORA EV & GWM Pickups in Iraq. (n.d.). GWM.iq. Retrieved March 1, 2026, from <https://gwm.iq/en>

- [23] BYD SEAL EV 2025 Flagship 2 motors. (n.d.). IQCars.net. Retrieved March 1, 2026, from <https://www.iqcars.net/en/carmodels/sfx/696>
- [24] Iraqi cars sales - Facts & data 2025. (2026). Focus2Move.com. Retrieved March 1, 2026, from <https://www.focus2move.com/iraqi-cars-sales/>
- [25] Iraq ICE & electric passenger vehicle market | Outlook. (n.d.). 6Wresearch. Retrieved March 1, 2026, from <https://www.6wresearch.com/industry-report/iraq-ice-electric-passenger-vehicle-market>
- [26] Iraq ICE & electric passenger vehicle market 2024-2030. (2024). ResearchAndMarkets.com. Retrieved March 1, 2026, from https://www.researchandmarkets.com/reports/5972642/iraq-ice-and-electric-passenger-vehicle-market?srsId=AfmBOorRyOviKQTDcrxeo_gJ4teF3AGOf-kILUMbxSq1PSwi8vEV7j3
- [27] Iraq new and used vehicle market. (n.d.). KenResearch.com. Retrieved March 1, 2026, from <https://www.kenresearch.com/industry-reports/iraq-new-28>. International Energy Agency (IEA). (2025). Global EV Outlook 2025: Trends and prospects in emerging economies.
- [28] Iraqi Kurdistan's EV revolution. (2025, September 23). ForeignPolicy.com. Retrieved March 1, (2026), from <https://foreignpolicy.com/2025/09/23/iraq-kurdistan-region-electric-auto-china-byd-economy-energy/>
- [29] Iraqi Customs Board. (2025). New tariff regulations for electric vehicles.
- [30] Cultural factors in Iraqi vehicle adoption. (2025). Iraqi Social Review.
- [31] 6Wresearch. (2025). Iraq Automotive Electric Vehicle Market (2025-2031): Trends, Outlook & Forecast (Report No. ETC10908944).
- [32] AlgoDriven. (2026). AI-Powered Vehicle Valuation in MENA Markets: Technology and Implementation. Dubai: AlgoDriven.
- [33] Kurdistan24. (2026, February 20). Iraq reimposes vehicle tariffs, sparks price hikes across markets. <https://www.kurdistan24.net/en/story/895342/>.