Erciyes University Journal of Institute Of Science and Technology

Research Article
10.65520/erciyesfen.2001917

Imprint:
Volume: 42(3)
Year: 2026
Page: 375-395

Hakan Onay[a]
Rukiye Nur Kaçmaz[b*]

[a] Undergraduate Student, Erciyes University, 1031510017@erciyes.edu.tr
[b] Asst. Prof., Erciyes University, rukiyekacmaz@erciyes.edu.tr

[*] Corresponding Author

Received: 7/24/2026
Accepted: 9/29/2026

Citation:
Hakan Onay, Rukiye Nur Kaçmaz, (2026). Comparing Global and Segment-Aware Machine Learning Models for Used Car Price Prediction. *Erciyes University Journal of Institute Of Science and Technology*, 42(3), 375-395. https://doi.org/10.65520/erciyesfen.2001917

Screened by iThenticate for Authors & Researchers

Except where otherwise noted, content in this article is licensed under a Creative Commons 4.0 International license. Icons by Font Awesome.

Comparing Global and Segment-Aware Machine Learning Models for Used Car Price Prediction

Abstract

Background: Used car prices are influenced by model year, mileage, fuel type, transmission, engine size, engine power and market position. This study compares global and segment-aware machine learning approaches for used car price prediction using the CarDekho dataset from Kaggle. Methods: After 167 exact duplicate extras were removed, 15,244 records were evaluated. The primary analysis used ten sealed 80/20 outer splits. Segment topology was selected by three-fold cross-validation within each outer-training set, and fixed CatBoost/XGBoost expert profiles were combined through an XGBoost routing classifier. Global, hard-routing, soft-gating and oracle variants were compared. All target models used log1p prices and a three-fold out-of-fold Duan retransformation correction. Results: Across the full-market splits, global XGBoost achieved R^2=0.8523±0.1120, RMSE=349,460±180,834 INR and MAE=104,138±5,274 INR. Soft gating achieved R^2=0.8733±0.1007, RMSE=321,661±173,497 INR and MAE=99,196±5,479 INR. Paired Wilcoxon p values were 0.0039 for R^2 and RMSE and 0.0020 for MAE. In the seed-42 paired bootstrap, the R^2 and RMSE intervals included zero, whereas the MAE reduction remained supported. The secondary IQR-restricted analysis yielded almost identical mean R^2 and RMSE for global and soft-gated models, with a lower MAE for soft gating. Conclusion: Segment-aware modeling reduced average full-market error, but the size of the advantage varied across splits. The contribution is a controlled empirical comparison rather than a new learning algorithm.

Keywords: used car price prediction, segment-aware modeling, mixture of experts, soft gating, XGBoost, nested validation.

İkinci El Araç Fiyat Tahmininde Genel ve Segment-Duyarlı Makine Öğrenmesi Modellerinin Karşılaştırılması

Öz

Arka plan: İkinci el araç fiyatları; model yılı, kilometre, yakıt türü, şanzıman, motor hacmi, motor gücü ve aracın pazardaki konumu gibi birçok etkene bağlıdır. Bu çalışmada CarDekho veri seti kullanılarak global ve segment duyarlı makine öğrenmesi yaklaşımları karşılaştırılmıştır. Yöntem: Tam yinelenen 167 ek kayıt çıkarıldıktan sonra 15.244 kayıt değerlendirilmiştir. Birincil analizde on adet kapalı %80/%20 dış veri bölümü kullanılmış; segment yapısı yalnızca her dış eğitim kümesindeki üç katlı çapraz doğrulamayla seçilmiştir. Önceden sabitlenen CatBoost/XGBoost uzmanları, XGBoost yönlendirme sınıflandırıcısıyla birleştirilmiş; global, hard routing, soft gating ve oracle yaklaşımları karşılaştırılmıştır. Log1p hedefi için üç katlı out-of-fold Duan geri dönüş düzeltmesi uygulanmıştır. Bulgular: Tam pazar bölümlerinde global XGBoost için R^2=0,8523±0,1120, RMSE=349.460±180.834 INR ve MAE=104.138±5.274 INR; soft gating için sırasıyla 0,8733±0,1007,

321.661±173.497 INR ve 99.196±5.479 INR bulunmuştur. Eşleştirilmiş Wilcoxon p değerleri R^2 ve RMSE için 0,0039, MAE için 0,0020'dir. Seed 42 bootstrap analizinde R^2 ve RMSE güven aralıkları sıfırı içerirken MAE azalması desteklenmiştir. İkincil IQR analizinde global ve soft-gated modellerin ortalama R^2 ve RMSE değerleri birbirine çok yakın, soft-gated modelin MAE değeri daha düşük bulunmuştur. Sonuç: Segment duyarlı modelleme tam pazarda ortalama hatayı azaltmıştır; ancak avantajın büyüklüğü veri bölümüne göre değişmektedir. Çalışmanın katkısı yeni bir algoritma değil, kontrollü bir ampirik karşılaştırmadır.

Anahtar kelimeler: ikinci el araç fiyat tahmini, segment duyarlı modelleme, uzmanlar karışımı, soft gating, XGBoost, nested doğrulama.

1. Introduction

1.1 Background and Motivation

The used car market has become increasingly important as buyers seek vehicles that fit their budgets and vehicles depreciate over time. A central concern for both buyers and sellers is determining a realistic selling price. Sellers may incur losses when a vehicle is priced below its market value, whereas buyers may overpay when the price exceeds the vehicle's actual market value. Accurate price estimation is therefore important for both parties. Used car prices are influenced by brand, model, age, mileage, engine size, engine power, fuel type, transmission type, number of seats and seller type. The effect of each feature may vary across vehicles; for example, mileage may be particularly influential for some models, whereas brand or model popularity may be more important for others. Traditional pricing relies on comparable listings, seller experience and market observation, but these sources may produce inconsistent estimates. Machine learning provides a data-driven alternative by evaluating multiple vehicle characteristics simultaneously. Used car price prediction is formulated as a regression problem because the target is a continuous selling price. The relationships between vehicle features and price are not necessarily linear, and vehicles in different price ranges may follow different pricing patterns.

This study examines whether a segment-aware approach can complement conventional global modeling in used car price prediction. Most prediction studies train a single model on the entire dataset, which may overlook differences among low-, mid- and high-priced vehicles. Age and mileage may be more influential in lower price ranges, whereas brand, model, engine power and transmission may have greater importance in higher price ranges. A segment-aware structure addresses this possibility by learning separate models for different price groups. The study therefore compares global regression models with price-segmented models and uses a routing classifier and soft gating to assign new vehicles to the relevant segment models and reduce abrupt prediction changes near segment boundaries.

1.2 Related Work

Early used-car studies mainly compared global prediction models. Pudaruth [1] evaluated linear regression, k-nearest neighbours, decision trees and Naive Bayes on a small Mauritian dataset. Monburinon et al. [2] later compared multiple linear regression, Random Forest and gradient-boosted trees on more than 300,000 German listings and found the boosted-tree model to be the strongest of the three. These studies show both the nonlinear nature of vehicle pricing and the importance of dataset size. Gegic et al. [3] combined artificial neural networks, support vector machines and Random Forest for vehicle-price classes. Recent studies have increasingly favored boosting and model fusion. Cui et al. [4] combined a residual network with an iterative XGBoost-LightGBM framework, while Huang et al. [5] reported R^2 = 0.9827 with XGBoost and R^2 = 0.9845 with a four-model fusion. These results are useful context, but they are not directly comparable because the datasets, target ranges, features and validation procedures differ. Tree ensembles can model nonlinear interactions without requiring a fixed linear form [6, 7]. XGBoost adds regularized boosting [8], while CatBoost addresses prediction bias and categorical features [9]. Deep tabular alternatives include TabNet [10] and FT-Transformer [11], although tree-based models often remain strong on medium-sized tabular datasets

CC: BY-NC-ND

[12]. The segment-aware design is related to mixture-of-experts models, in which specialized regressors act as experts and a gating model assigns inputs to one or more experts [13, 14]. KMeans can be used to define price regimes from the training distribution [15]. This study therefore compares global models with target-derived segments, learned routing and probability-weighted soft gating.

Recent research has also emphasized feature detail, preprocessing, model combination, uncertainty and evaluation design. Surjandari et al. [16] combined stacked generalization with sequential model-based optimization and identified initial vehicle price as an important determinant of residual value. Madhusudhanan et al. [17] proposed ProbSAINT, an attention-based probabilistic tabular model that provides both point predictions and uncertainty estimates. Using 92,239 real-world sales, Bergmann and Feuerriegel [18] found that granular vehicle-equipment information improved MAE by 3.27%. Fayyaz et al. [19] likewise reported substantial performance gains after feature engineering across linear, ensemble and neural models. More recent studies extended these directions: Chen et al. [20] combined CatBoost with BiLSTM for residual-value prediction of secondhand new-energy vehicles, while Netayawijit et al. [21] benchmarked multiple train-test ratios with leakage-free preprocessing and found ensemble methods more accurate than Linear Regression.

1.3 Contributions of the Study

The main contribution of this study is the evaluation of a segment-aware structure alongside a single global model for used car price prediction. The revised analysis does not present segmentation or mixture-of-experts modeling as a new algorithm. Instead, it tests the practical value of this established idea under repeated, leakage-controlled evaluation on the same dataset.

Three segmentation strategies were examined: a two-segment quantile approach, a three-segment quantile approach and a KMeans-based three-segment approach. The topology was selected separately inside each outer-training set, allowing fixed-proportion and distribution-derived price groups to be compared without using the outer test outcomes.

During testing, a routing classifier estimated the price segment of each new vehicle. Hard routing selected one expert, whereas soft gating combined all expert predictions according to the estimated segment probabilities. An oracle route was included only as an unattainable diagnostic upper bound. Repeated splits, paired tests, router diagnostics, duplicate checks, group-aware splitting and subgroup errors were used to show where any gain arose and how stable it was.

2. Materials and Methods

2.1 Dataset

This study used the open-source CarDekho dataset published on Kaggle [22]. The original dataset contained 15,411 records and 14 columns. The identifier-like columns Unnamed: 0 and car_name were removed from modeling, and 167 exact duplicate extras were eliminated before splitting. The resulting analysis data contained 15,244 records, 11 predictors and the selling_price target. The records represent the Indian market, and no currency conversion was applied; therefore, selling_price and all error metrics are reported in Indian rupees (INR). Numerical variables included vehicle age, distance driven, mileage, engine size, maximum power and number of seats. Categorical variables included brand, model, seller type, fuel type and transmission type.

The target variable was continuous, and the task was therefore formulated as a regression problem. The full-market analysis retained the complete price range and was treated as the primary evaluation. A secondary IQR-restricted analysis used limits calculated only from the training prices after each outer split. Numerical and categorical information was processed together because vehicles with similar mileage may still differ substantially in brand, model, engine, fuel type or transmission.

The duplicate audit identified 329 records belonging to exact duplicate sets. After exact de-duplication, 953 groups (2,202 records) had identical modeling predictors but more than one observed selling price. These records were retained because the price differences may reflect unrecorded condition, location or timing rather than literal duplication. Their possible dependence

CC: BY-NC-ND

was examined separately with brand-model-vehicle_age group-aware splits.

2.2 Data Preprocessing

The dataset required several preprocessing steps. Columns that were not useful for prediction were removed, exact duplicate rows were collapsed, and missing values were checked; the supplied dataset contained no incomplete rows. The selling_price distribution had a long upper tail. These values were retained in the primary full-market analysis. For the secondary IQR-restricted analysis, the IQR limits were learned only from the training prices. A model category occurring fewer than 25 times in the current training fold was grouped under Other. This mapping, OneHotEncoder(handle_unknown='ignore') and StandardScaler were fitted inside the corresponding training pipeline. The test ratio was 20%.

In the representative seed-42 split, the training-price quartiles were Q1=380,000 and Q3=815,500 INR. The nominal 1.5×IQR fences were −273,250 and 1,468,750 INR; because the lower fence was negative, the effective lower boundary was the observed market minimum of 40,000 INR. Applying the training-defined range retained 11,086 training and 2,766 test records. Test membership in this cohort still depends on the true test price, so the IQR results are conditional rather than deployment-realistic.

Categorical and numerical variables were processed separately within each model pipeline. OneHotEncoder was applied to brand, model, seller_type, fuel_type and transmission_type, creating binary indicator columns for the observed categories.

StandardScaler was applied to the numerical variables because their ranges differed substantially. For example, km_driven can take much larger values than seats. Scaling placed the numerical variables on more comparable scales for models that are sensitive to feature magnitude.

ColumnTransformer combined the categorical and numerical preprocessing steps. OneHotEncoder was applied to categorical variables and StandardScaler to numerical variables. All fitted statistics and category mappings were learned from the training data used in each fit [23].

The revised primary evaluation used ten 80/20 train-test partitions with random seeds 11, 23, 37, 42, 59, 71, 83, 97, 109 and 131. After exact de-duplication, every full-market split contained 12,195 training and 3,049 test records. Each outer test partition remained sealed until preprocessing, topology selection, expert fitting, router probability selection and the Duan factor had been completed using the outer-training data.

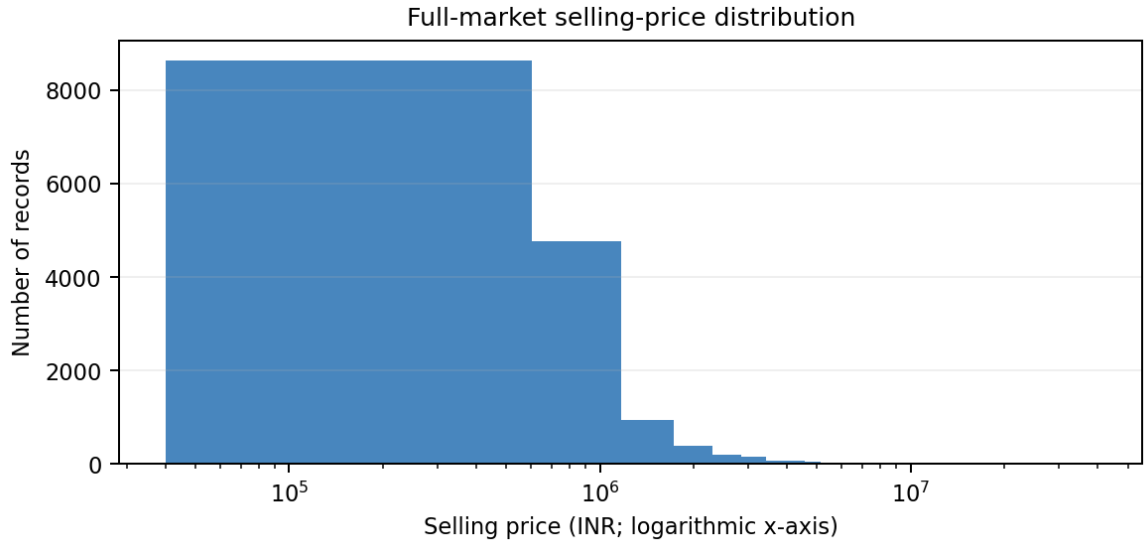

Figure 1. *Full-market selling-price distribution. A logarithmic horizontal axis displays the long upper tail without removing deployment-relevant records.*

CC: BY-NC-ND

Figure 1 shows the strongly right-skewed full-market distribution. The tail was retained for the primary analysis, while log-transformed prices were used during model fitting to reduce its influence on optimization.

2.3 General Procedure

Two main regression frameworks were evaluated: global and segment-aware modeling. Within each evaluation cohort, the global XGBoost model learned one relationship across vehicles. The segment-aware framework divided only the training targets into price groups, fitted registered expert profiles within those groups and used a feature-based router for new vehicles. Global, hard-routing, soft-gating and oracle predictions were calculated on the same sealed test rows and compared using R^2, RMSE and MAE.

2.4 Global Regression Models

In the global modeling framework, one model was trained to learn a common relationship between the vehicle features and selling price across low-, mid- and high-priced cars. This approach provides a simple benchmark but does not explicitly model differences among price groups.

The original global-model screening included Linear Regression, Ridge Regression and Lasso Regression as linear methods; Random Forest, Gradient Boosting, XGBoost and CatBoost as tree-based or boosting methods; and TabNet, FT-Transformer and SSL-AutoEncoder as deep tabular learning methods. XGBoost was retained as the fixed global baseline for the revised repeated analysis, while CatBoost and XGBoost served as the registered segment experts.

Random Forest, Gradient Boosting, XGBoost and CatBoost were evaluated because tree-based methods can capture nonlinear relationships and interactions among mixed vehicle characteristics. CatBoost received the same one-hot encoded input as the other estimators in this implementation; native categorical handling was not used. The exact settings of the revised models and the retained exploratory models are listed in Appendix A.

Linear Regression assumes a linear relationship between the inputs and target. Ridge and Lasso are regularized variants of linear regression [24, 25]. Ridge reduces coefficient magnitudes to control overfitting, whereas Lasso may shrink some coefficients to zero and therefore produce a feature-selection effect. Their linear structure may nevertheless limit their ability to represent complex price relationships.

Random Forest combines predictions from multiple decision trees, whereas Gradient Boosting trains weak learners sequentially to reduce residual errors. XGBoost is a regularized and optimized gradient-boosting implementation, while CatBoost is designed to handle categorical features and reduce prediction bias. These models are well suited to nonlinear interactions among variables such as mileage, age, brand and engine power.

TabNet uses attentive feature selection, FT-Transformer applies a transformer architecture to tabular data, and SSL-AutoEncoder first learns an unlabeled representation before regression fine-tuning. The legacy TabNet profile used n_d=n_a=16, three decision steps, gamma=1.3 and a learning rate of 10^{-4}. FT-Transformer used a token size of 32, four heads, three blocks and dropout of 0.15. Both used batches of 256, at most 30 epochs and early-stopping patience of 6.

The SSL-AutoEncoder first reconstructed the processed input without using price labels and then fine-tuned a regression head for log-price, following the representation-learning idea of denoising autoencoders [26]. Its encoder used 128 hidden units and a 64-dimensional latent representation. It was pretrained for at most 20 epochs and fine-tuned for at most 30 epochs using AdamW, a learning rate of 5×10^{-4}, batch size 256, dropout 0.15 and patience 6.

An audit of the original implementation showed that the deep candidates did not share an equivalent hyperparameter search, compute budget and nested validation procedure. The extremely negative TabNet score was therefore not treated as a comparative finding, and all deep-model scores were removed from the revised inferential comparison. Their architectures are disclosed for

CC: BY-NC-ND

provenance only; a fair deep-learning benchmark would require matched search spaces, compute, repeated seeds and the same sealed outer protocol.

2.5 Log Transformation

The selling_price variable covered a wide range, with a small number of vehicles priced substantially higher than most observations. A logarithmic transformation was therefore applied during training to reduce skewness and make the target distribution easier to model.

Models were trained on z=log(1+selling_price). Direct inversion with exp(ẑ)−1 can underestimate the conditional mean when log-scale residual variance is non-zero. For every fitted training set, three-fold out-of-fold log predictions were therefore used to estimate Duan's smearing factor [27]. The correction was learned without test residuals and final R^2, RMSE and MAE were calculated only after predictions had been returned to the INR scale.

$$S = n^{-1} \Sigma_i \exp[\log(1+y_i) - \hat{z}_i], \quad \hat{y} = S[\exp(\hat{z}) - 1].$$

For the seed-42 global model, S=1.01437. Relative to direct inversion, the correction reduced systematic underprediction from approximately 12,126 INR to about 1,200 INR. This adjustment concerns retransformation bias; it does not use or optimize the test errors.

2.6 Segment-Aware Framework

Global models learn one relationship across all vehicles, whereas segment-aware models learn separate relationships within different price groups. The approach was motivated by the possibility that age and mileage may be more influential in lower price ranges, while brand, model, engine power and transmission may have greater influence in higher price ranges.

The segment-aware framework consisted of three stages. First, a segment topology was selected inside the outer-training set by three-fold cross-validation. Second, the registered expert profiles were fitted within the selected training-price groups: CatBoost for low and middle segments and XGBoost for the high segment; the two-segment profile used CatBoost for low and XGBoost for high. Third, a routing classifier estimated segment membership for each test vehicle from its observed features. The same expert predictions supported hard routing, soft gating and the oracle diagnostic.

2.6.1 Segmentation Strategies

The 2-Segment Quantile strategy divided training vehicles at the 40th percentile of selling_price. The lower group was labeled Low and the remaining group High. The boundary was re-estimated in every outer-training split.

The 3-Segment Quantile strategy divided the training vehicles into Low, Mid and High groups at the 33rd and 67th percentiles. The groups were approximately balanced before router down-sampling. The name 3-Segment Quantile is used consistently throughout the revised manuscript.

The 3-Segment KMeans strategy applied KMeans with three clusters and n_init=20 to log-transformed training prices. The resulting clusters were ordered from low to high price. Unlike quantile segmentation, this approach defined groups from the distributional structure of the training target. KMeans was refitted separately within every training or cross-validation fold.

2.6.2 Routing Classifier

Price-based segments can be defined during training because the selling price is known. During testing, however, the actual selling price is unavailable. A routing classifier was therefore required to estimate the price segment of each new vehicle from its observed characteristics.

The routing classifier predicted segment membership from vehicle features and directed each vehicle to the corresponding expert models. For example, a vehicle assigned a high probability of belonging to the high-price segment received a larger contribution from the high-price regression model.

CC: BY-NC-ND

The revised router was an XGBoost classifier with 250 trees, learning_rate=0.05, max_depth=5, subsample=0.90 and colsample_bytree=0.90. Router training was balanced by down-sampling each class to the smallest training segment. Within each outer-training set, raw probabilities were compared with three-fold sigmoid calibration on an internal stratified 20% validation subset; the method with lower validation log loss was refitted on all outer-training records. Only vehicle features were used to generate test-time probabilities.

2.6.3 Soft Gating

Hard routing assigns each vehicle to a single segment model. This may increase errors near segment boundaries when the routing classifier is uncertain or assigns a vehicle to the wrong group.

Soft gating was used to reduce dependence on a single routing decision. Let $f_k(x)$ denote the prediction of expert k and $p_k(x)$ the router probability for that expert. Hard routing selects the most probable expert, whereas soft gating calculates the probability-weighted mixture shown below, following the mixture-of-experts principle [13, 14].

Hard: $k^* = \arg\max_k p_k(x), \hat{y}(x)=f_{k^*}(x)$; *Soft:* $\hat{y}(x)=\Sigma_k p_k(x)f_k(x), \Sigma_k p_k(x)=1.$

Soft gating is particularly relevant for vehicles with characteristics close to the boundary between two price groups. Rather than relying entirely on one expert model, the final estimate incorporates the predictions of multiple segment models according to the router probabilities.

The revised analysis reports global XGBoost, hard routing, soft gating and oracle routing on identical outer test records. Oracle routing uses the true training-defined test segment only to select among already fitted experts and is not a deployable method. The hard-versus-soft contrast isolates the effect of probability-weighted blending, while the soft-versus-oracle gap indicates the remaining cost of imperfect routing.

2.7 Experimental Protocol and Reproducibility

The primary full-market analysis used ten sealed outer test partitions rather than one favorable random split. Within each outer-training set, three-fold cross-validation compared the three candidate segment topologies using validation R^2. The winning topology was then refitted on the full outer-training set with expert profiles fixed before the test set was opened. Cross-validation scores used for selection were calculated on held-out training folds; final outer-test R^2, RMSE and MAE were calculated on the original INR scale.

Rare-model grouping, OneHotEncoder, StandardScaler, quantile boundaries, KMeans, Duan factors and the router probability method were fitted or selected using training data only. The full-market metrics were calculated without target-based test exclusion. The IQR-restricted cohort used training-derived fences but still required the true test price to determine cohort membership, so it was interpreted as a secondary conditional analysis. A further sensitivity check used five GroupShuffleSplit partitions with brand-model-vehicle_age as the group identifier and verified zero group overlap in every fold.

Across the ten outer splits, results are summarized by the mean, sample standard deviation and 95% t confidence interval. Paired two-sided Wilcoxon signed-rank tests compared hard, soft and oracle predictions with global XGBoost. For seed 42, 5,000 paired bootstrap resamples of the same test rows provided confidence intervals for soft-minus-global differences. Router assessment included accuracy, balanced accuracy, macro and class-specific F1, multiclass Brier score, log loss, expected calibration error, a confusion matrix and one-vs-rest reliability curves.

The validated run used Python 3.12.13, scikit-learn 1.6.1, XGBoost 3.0.5 and CatBoost 1.2.8 on CPU. Mean time per full-market outer split was 9.72 s for topology selection, 0.88 s for the global model and 6.64 s for the final experts and router. Appendix A gives the complete settings, software versions, hardware and legacy-model audit.

CC: BY-NC-ND

2.8 Evaluation Metrics

Performance was evaluated using R^2, RMSE and MAE. R^2 measures the proportion of target variation explained by a model, with values closer to 1 indicating a better fit. RMSE gives larger errors more weight, whereas MAE is the average absolute difference between the observed and predicted prices. All final errors were reported in INR, and lower RMSE and MAE indicate better performance. The definitions used were $R^2=1-\Sigma(y_i-\hat{y}_i)^2/\Sigma(y_i-\bar{y})^2$, $RMSE=\sqrt{[n^{-1}\Sigma(y_i-\hat{y}_i)^2]}$ and $MAE=n^{-1}\Sigma|y_i-\hat{y}_i|$.

2.9 Ethical Considerations

This study used a publicly available secondary dataset containing vehicle characteristics and prices. The dataset did not include personally identifiable human data or animal participant data; therefore, ethics committee approval and informed consent were not required.

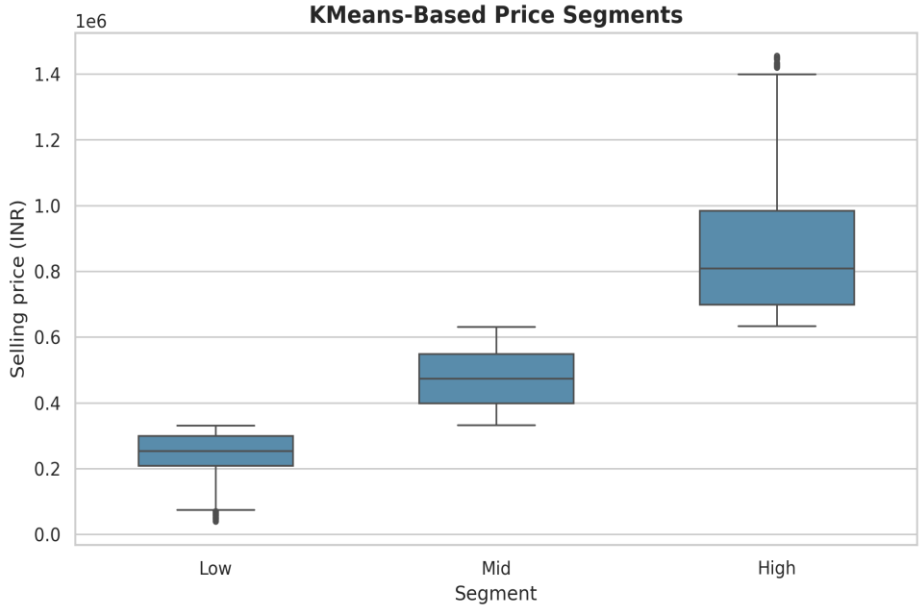

Figure 2. *Illustrative KMeans-based training-price segments ordered from low to high (INR).*

Figure 2 illustrates how the KMeans candidate formed low-, middle- and high-price groups from one training partition. In the revised protocol, this clustering was repeated inside the relevant training fold and never fitted to a sealed test target.

3. Results

The primary estimand in this revision is full-market performance, summarized at the start of the Results and reported in detail in Section 3.3. Across ten paired full-market splits, soft gating reduced mean RMSE by 27,798 INR and mean MAE by 4,941 INR relative to global XGBoost. Sections 3.1 and 3.2 retain the original IQR-based analytical sequence as a secondary continuity analysis.

3.1 Global Model Results in the IQR-Restricted Cohort

The initial seed-42 model screen in the IQR-restricted cohort identified XGBoost as the strongest conventional global model, with test R^2=0.8945, RMSE=90,713 INR and MAE=64,946 INR. Random Forest, CatBoost and Gradient Boosting also produced strong descriptive results. This screen was retained because it explains the choice of XGBoost as the fixed global baseline, but it is not the basis of the repeated inferential comparison.

Among the retained tree ensembles, Random Forest achieved R^2=0.8906, CatBoost 0.8852 and Gradient Boosting 0.8819. Their performance supports the use of nonlinear models for mixed vehicle attributes, where mileage or engine power may have different effects across brands, ages and fuel types.

CC: BY-NC-ND

Linear Regression and Ridge Regression achieved R^2 values of 0.8672 and 0.8658, respectively, whereas Lasso achieved 0.6814. The difference between the regularized linear models and the tree ensembles is consistent with interactions and nonlinearities that a single linear function does not represent easily.

Deep-model scores are not shown in Table 1. Their original training and validation budgets were not sufficiently equivalent for a fair comparison, and the unstable negative TabNet result was not interpreted as evidence against TabNet. The revised claims therefore rely on the registered XGBoost/CatBoost profiles evaluated in the common repeated protocol.

Table 1. Descriptive Global Model Screening in the Seed-42 IQR-Restricted Test Cohort

Model	**Train R^2 (log)**	**5-fold CV R^2 mean ± SD (log)**	**Test R^2 (INR scale)**	**RMSE (INR)**	**MAE (INR)**
XGBoost	0.9203	0.9013 ± 0.0029	0.8945	90,713	64,946
Random Forest	0.9586	0.8960 ± 0.0029	0.8906	92,374	65,549
CatBoost	0.9059	0.8968 ± 0.0033	0.8852	94,628	67,203
Gradient Boosting	0.9076	0.8945 ± 0.0026	0.8819	95,987	68,493
Linear Regression	0.8875	0.8842 ± 0.0041	0.8672	101,771	71,635
Ridge Regression	0.8872	0.8841 ± 0.0041	0.8658	102,339	71,981
Lasso Regression	0.7985	0.7976 ± 0.0063	0.6814	157,659	104,539

3.2 Segment-Aware Model Results in the IQR-Restricted Cohort

The secondary IQR analysis compared global XGBoost with hard routing, soft gating and oracle routing across the same ten outer seeds. Segment topology was chosen within each outer-training cohort rather than fixed after observing test performance. Table 2 reports mean±SD and 95% confidence intervals for every method.

Global XGBoost achieved mean R^2=0.8928±0.0026, RMSE=91,092±1,683 INR and MAE=65,517±961 INR. Soft gating produced nearly the same R^2 (0.8928±0.0026) and RMSE (91,072±1,489 INR), while reducing mean MAE to 64,933±930 INR.

Hard routing was weaker in this restricted cohort, with R^2=0.8810±0.0057, RMSE=95,932±1,966 INR and MAE=68,203±1,221 INR. Oracle routing reached R^2=0.9204±0.0110 and MAE=54,833±4,159 INR, indicating that much of the remaining headroom was related to segment assignment rather than the absence of distinct expert predictions.

For the primary full-market runs, training-only selection chose 3-Segment Quantile in eight splits and 3-Segment KMeans in two; 2-Segment Quantile was evaluated but not selected. In seed 42, the selected 3-Segment Quantile boundaries were ≤449,000 INR, 449,000–700,000 INR and >700,000 INR. The Low, Mid and High groups contained 4,068, 4,124 and 4,003 training records and 999, 1,009 and 1,041 test records, respectively. Their registered experts were CatBoost, CatBoost and XGBoost.

CC: BY-NC-ND

Table 2. Global and Segment-Aware Results in the Secondary IQR-Restricted Cohort Across Ten Splits

Approach	**Routing**	**n**	**R^2 mean ± SD [95% CI]**	**RMSE mean ± SD [95% CI]**	**MAE mean ± SD [95% CI]**
Global XGBoost	none	10	0.8928 ± 0.0026 [0.8909, 0.8946]	91,092 ± 1,683 [89,888, 92,296]	65,517 ± 961 [64,830, 66,204]
Hard routing	hard	10	0.8810 ± 0.0057 [0.8769, 0.8851]	95,932 ± 1,966 [94,525, 97,339]	68,203 ± 1,221 [67,329, 69,076]
Soft gating	soft	10	0.8928 ± 0.0026 [0.8909, 0.8947]	91,072 ± 1,489 [90,007, 92,137]	64,933 ± 930 [64,267, 65,598]
Oracle routing	true segment diagnostic	10	0.9204 ± 0.0110 [0.9125, 0.9282]	78,344 ± 5,796 [74,198, 82,491]	54,833 ± 4,159 [51,857, 57,808]

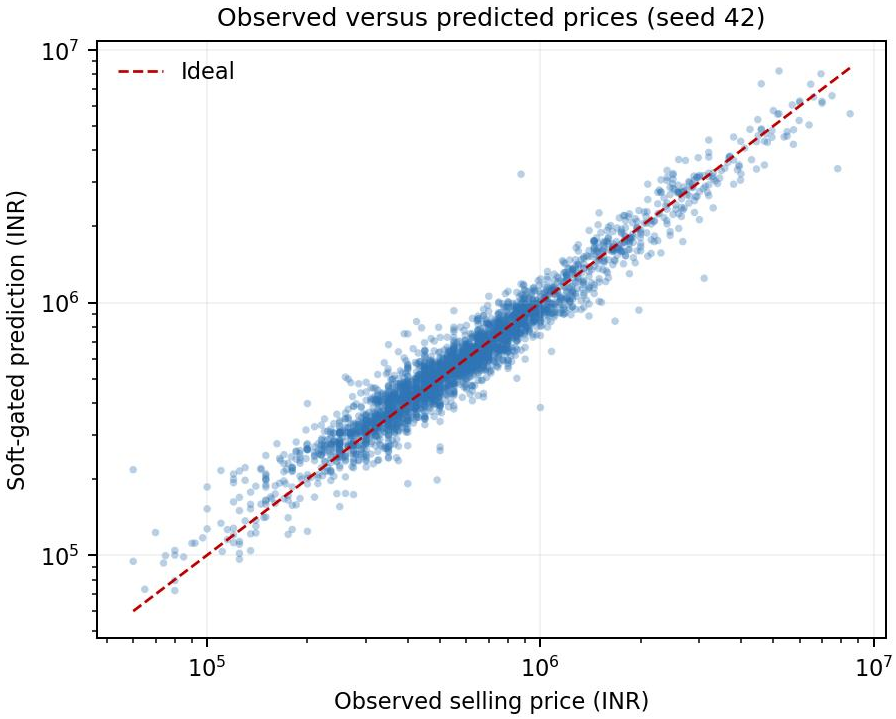

Figure 3*. Observed versus soft-gated predicted selling prices for the seed-42 full-market test set. Both axes are logarithmic for display; all reported metrics use untransformed INR.*

Figure 3 shows that most predictions followed the observed-price diagonal, while the largest departures occurred in the sparse high-price tail. This tail also explains why RMSE varied much more across outer splits than MAE.

CC: BY-NC-ND

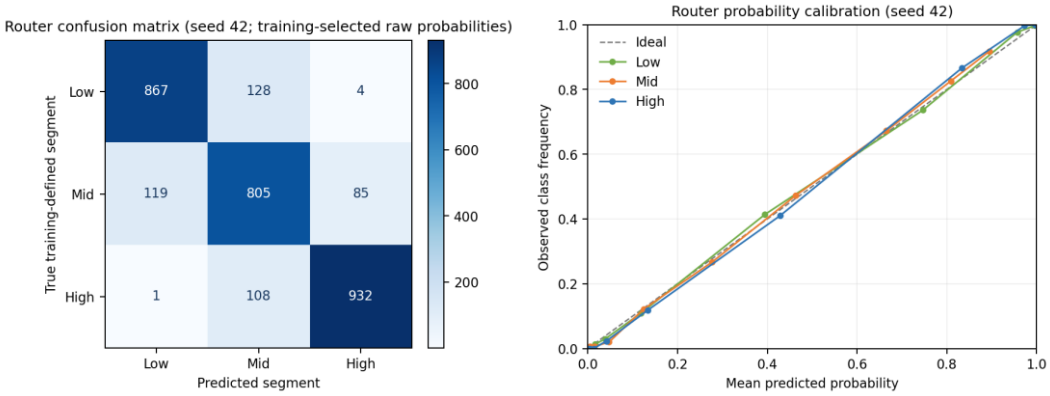

***Figure 4**. Seed-42 router diagnostics: (a) confusion matrix and (b) one-vs-rest reliability curves for the training-selected raw probability method.*

The confusion matrix in Figure 4 shows that most errors occurred between neighboring price regimes. The reliability curves are diagnostic rather than evidence that post-hoc calibration must always be used; the probability method was selected anew from training data in each outer split.

In seed 42, global XGBoost achieved R^2=0.9256, RMSE=212,737 INR and MAE=99,800 INR, whereas soft gating achieved R^2=0.9312, RMSE=204,577 INR and MAE=94,642 INR. In 5,000 paired bootstrap resamples, the soft-minus-global R^2 difference was 0.0056 (95% CI −0.0065 to 0.0177; p=0.348), the RMSE difference was −8,160 INR (95% CI −26,526 to 8,880; p=0.348), and the MAE difference was −5,158 INR (95% CI −7,610 to −2,707; p<0.0002). Thus only the seed-42 MAE difference was clearly resolved.

The seed-42 router achieved accuracy=0.8541, balanced accuracy=0.8537 and macro-F1=0.8542. Low, Mid and High class F1 values were 0.8731, 0.7854 and 0.9040 on 999, 1,009 and 1,041 test records. The corresponding precision/recall pairs were 0.8784/0.8679, 0.7733/0.7978 and 0.9128/0.8953.

Probability quality was also examined because soft gating uses the probability magnitudes, not only the predicted class. The seed-42 multiclass Brier score was 0.2083, log loss 0.3457 and 10-bin expected calibration error 0.0180. Raw probabilities were retained because their training-only validation log loss (0.3499) was lower than that of sigmoid calibration (0.3788).

3.3 Full-Market Evaluation

The full-market analysis is the primary evaluation because it retained the complete deployment-relevant price range and did not use the true test target for inclusion. Across ten sealed outer splits, global XGBoost achieved R^2=0.8523±0.1120, RMSE=349,460±180,834 INR and MAE=104,138±5,274 INR. Hard routing yielded R^2=0.8718±0.1004, RMSE=323,965±172,699 INR and MAE=103,459±5,527 INR. Soft gating yielded R^2=0.8733±0.1007, RMSE=321,661±173,497 INR and MAE=99,196±5,479 INR. Oracle routing reached R^2=0.8776±0.1015, RMSE=314,973±175,886 INR and MAE=86,820±5,723 INR. Relative to global XGBoost, soft gating increased mean R^2 by 0.0210 and reduced RMSE by 27,798 INR and MAE by 4,941 INR. Paired Wilcoxon p values were 0.0039, 0.0039 and 0.0020, respectively. However, the relatively large between-split standard deviation in full-market RMSE should be emphasized: the mean reduction does not imply stable error magnitude in every partition and remains sensitive to the distribution of rare high-price vehicles.

CC: BY-NC-ND

Table 3. Primary Full-Market Performance Across Ten Sealed Outer Splits

Approach	**R^2 mean ± SD [95% CI]**	**RMSE mean ± SD [95% CI] (INR)**	**MAE mean ± SD [95% CI] (INR)**
Global XGBoost	0.8523 ± 0.1120 [0.7722, 0.9325]	349,460 ± 180,834 [220,099, 478,820]	104,138 ± 5,274 [100,365, 107,910]
Hard routing	0.8718 ± 0.1004 [0.8000, 0.9436]	323,965 ± 172,699 [200,424, 447,506]	103,459 ± 5,527 [99,505, 107,412]
Soft gating	0.8733 ± 0.1007 [0.8013, 0.9453]	321,661 ± 173,497 [197,549, 445,774]	99,196 ± 5,479 [95,277, 103,116]
Oracle routing	0.8776 ± 0.1015 [0.8050, 0.9502]	314,973 ± 175,886 [189,152, 440,794]	86,820 ± 5,723 [82,726, 90,914]

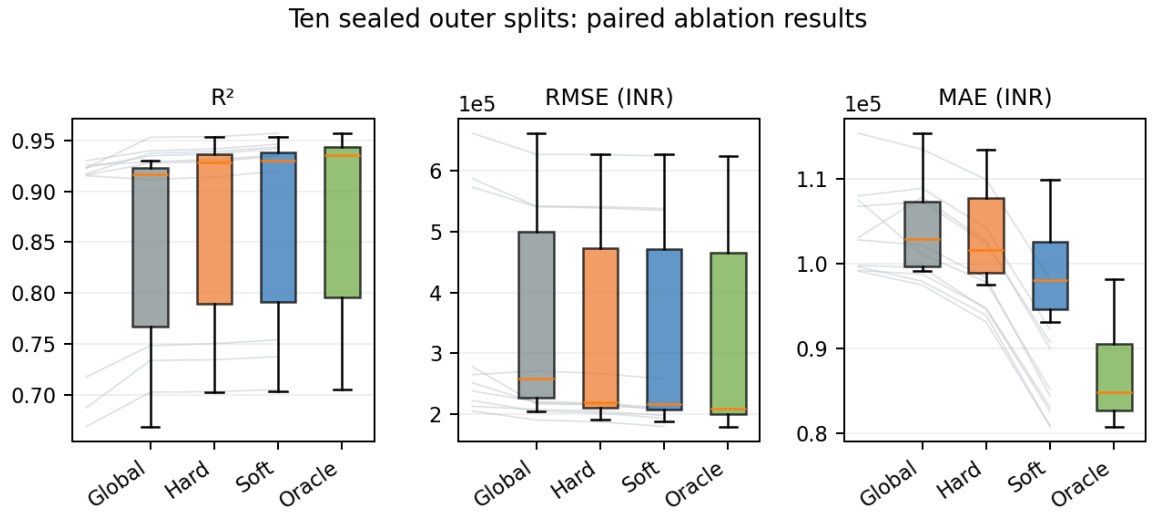

Figure 5. *Paired full-market ablation results across ten sealed outer splits. Lines join methods evaluated on the same test partition.*

Figure 5 makes the between-split variability visible. The same difficult high-price cases affected all four methods within a split, which is why paired tests were used instead of treating the method scores as independent samples.

A brand-model-vehicle_age group-aware sensitivity analysis used five additional splits with zero group overlap. Global XGBoost achieved mean R^2=0.8998, RMSE=258,865 INR and MAE=115,975 INR; soft gating achieved R^2=0.9092, RMSE=244,983 INR and MAE=110,345 INR. The direction of change therefore persisted when closely related vehicle families could not occur in both train and test sets.

CC: BY-NC-ND

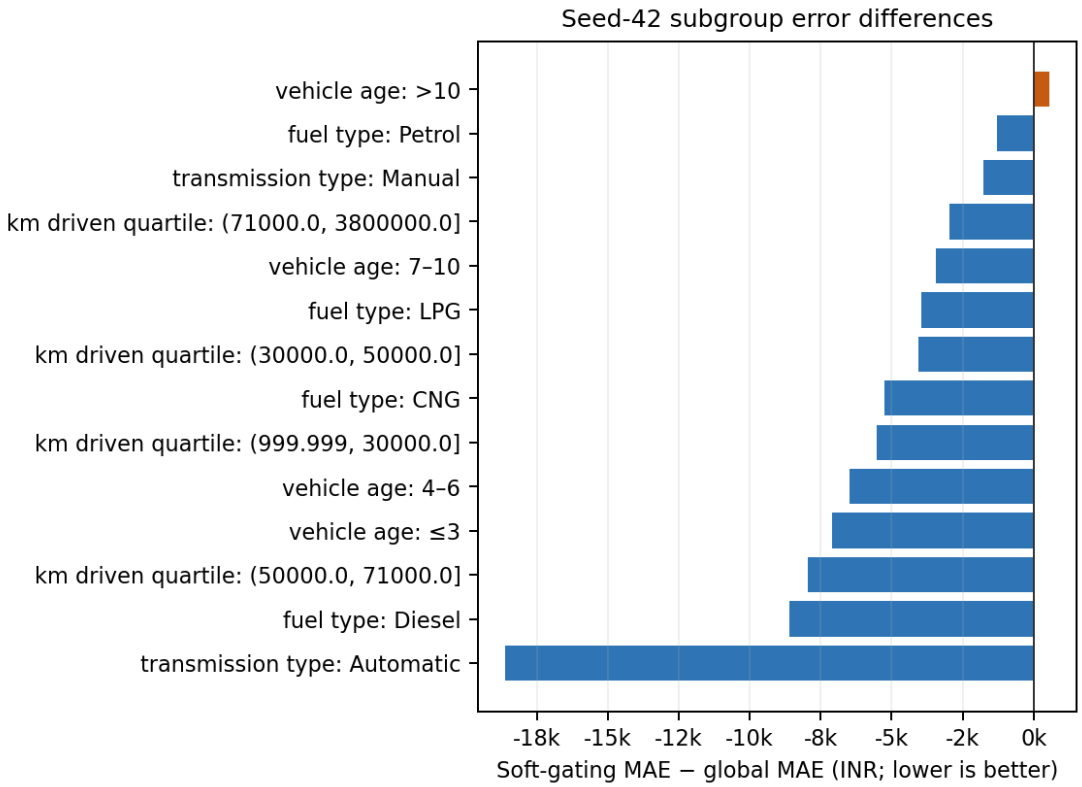

Figure 6. *Seed-42 subgroup MAE differences. Negative values favor soft gating; positive values favor global XGBoost.*

Subgroup errors were calculated by true price segment, vehicle-age band, mileage quartile, brand, fuel type and transmission. Soft gating reduced MAE in all displayed mileage, fuel and transmission groups and in three of four age groups; vehicles older than 10 years showed a small 564 INR increase. Among the ten most frequent named brands, the change favored soft gating for nine, while Volkswagen increased by 3,117 INR. These are descriptive seed-42 results, and small groups such as LPG vehicles should not support broad fairness claims.

4. Discussion

4.1 Global and Segment-Aware Results

The global results indicate that used car pricing is not adequately represented by a purely linear relationship. Linear Regression and Ridge Regression performed reasonably well but remained below the strongest tree-based models, which can capture nonlinear interactions among vehicle characteristics.

The performance of tree-based models can be attributed to their ability to learn conditional relationships. For example, the effect of mileage may vary according to vehicle age, brand or fuel type. Such interactions are difficult to represent with a single linear function.

In the primary full-market evaluation, global XGBoost remained a strong baseline, but soft gating improved the average R^2, RMSE and MAE across the ten paired splits. Importantly, the substantial between-split variability in full-market RMSE limits the strength and practical certainty of the average improvement. This variability was largely associated with differences in the composition of the sparse high-price tail across the outer test partitions, where a small number of difficult cases can exert a disproportionate influence on a squared-error metric. Therefore, the lower mean RMSE of soft gating should be interpreted as a modest empirical tendency within the present dataset rather than evidence of stable or uniform superiority across all splits. The paired design improves the reliability of the

CC: BY-NC-ND

within-dataset comparison, but repeated internal splitting cannot reproduce the distribution shifts that may occur across countries, sales platforms or market periods. Because no independent external dataset or chronologically separated test cohort was available, the geographic generalizability and temporal stability of the observed improvement remain unconfirmed and should be evaluated before any deployment-oriented claims are made.

This pattern is consistent with Monburinon et al. [2], who found gradient-boosted trees to be stronger than multiple linear regression and Random Forest on a large used-car dataset. It also agrees with Grinsztajn et al. [12], who showed that tree-based models remain difficult to surpass on many medium-sized tabular datasets.

The deep tabular models were retained in the methodological record but removed from the revised inferential comparison. Their original configurations did not receive a tuning and validation budget demonstrably equivalent to the tree ensembles, and the unstable TabNet run was not a reliable comparative result. Consequently, the revision makes no claim that the deep models are generally weaker or stronger.

The segment-aware framework was motivated by the possibility that pricing relationships differ across price ranges. Age and mileage may play a larger role among lower-priced vehicles, while brand, model, engine power and transmission may be more influential among higher-priced vehicles.

The ablation helps separate the components of the segment-aware framework. Hard routing improved mean full-market R^2 and RMSE but changed MAE only slightly relative to global XGBoost. Soft gating further reduced RMSE and, more clearly, MAE. Oracle routing performed best, showing that routing uncertainty still limited the fitted experts. Oracle performance is diagnostic rather than deployable because it uses the true test segment.

The strength of XGBoost is consistent with Monburinon et al. [2] and with the broader tabular benchmark of Grinsztajn et al. [12]. Results reported by Huang et al. [5] and Cui et al. [4] provide useful context for boosting and model fusion, but direct numerical comparison remains inappropriate because the datasets, target ranges, currencies, features and validation procedures differ.

The secondary IQR-restricted cohort produced higher and much more stable R^2 values than the full market. Within that restricted range, global and soft-gated R^2 and RMSE were nearly identical, although soft gating reduced MAE by approximately 584 INR on average. This contrast confirms that the reported advantage depends on the evaluated price range and that the full-market results should carry the main interpretation.

The recent literature also helps interpret the remaining error and the scope of the present findings. The benefits of feature engineering reported by Fayyaz et al. [19] and the strong ensemble results of Netayawijit et al. [21] are consistent with the performance of XGBoost in this study. Bergmann and Feuerriegel [18] and Surjandari et al. [16] further indicate that equipment details and initial vehicle price can add information not available in the CarDekho variables used here. ProbSAINT demonstrates the value of uncertainty-aware prediction beyond point-error metrics [17], while the CatBoost-BiLSTM model of Chen et al. [20] suggests that hybrid models may be useful for specialized and temporally sensitive vehicle markets. Direct numerical comparison remains inappropriate because these studies use different datasets, markets, variables and validation designs.

The effectiveness of a segment-aware framework depends on the segment definitions, expert fit, routing discrimination and probability quality. Most seed-42 routing errors occurred between adjacent price groups. Balanced accuracy and macro-F1 were close to ordinary accuracy, but the middle segment remained more difficult than the low and high segments.

The 3-Segment Quantile topology was chosen by training-only cross-validation in eight of ten full-market splits, while 3-Segment KMeans was chosen in two. This frequency supports the usefulness of approximately balanced price groups in this dataset without implying that one topology is universally best.

CC: BY-NC-ND

KMeans remained a viable candidate and was selected in the two splits where its training-only validation score was highest. Its less balanced price clusters may better represent distributional modes in some partitions, but they can also leave fewer observations for a high-price expert.

Overall, segmentation performance depended on the boundaries, selected expert profiles, router accuracy, probability calibration and evaluation metric. Re-estimating these elements within training data prevented a favorable test split from determining the reported architecture.

The repeated-split Wilcoxon tests supported the average soft-gating improvement, but the seed-42 bootstrap intervals for R^2 and RMSE included zero. The evidence therefore supports a modest average benefit, especially for MAE, rather than a large improvement that is guaranteed in every split.

4.2 Soft Gating

Hard routing assigns each vehicle to one segment model and can therefore be sensitive to incorrect routing decisions near segment boundaries.

Soft gating reduces this dependence by combining the predictions of all segment models according to their estimated probabilities. The final prediction is therefore not determined by a single expert.

The direct hard-versus-soft comparison indicates that probability-weighted blending contributed beyond segmentation and expert fitting. Across full-market splits, soft gating reduced mean RMSE by 2,304 INR and MAE by 4,263 INR relative to hard routing. The remaining gap to oracle routing shows that better segment inference or boundary-aware experts may still improve the system.

4.3 Strengths of the Study

A strength of the revised study is that the original model comparison is now supplemented by ten paired outer splits with sealed test sets. The segment topology and probability treatment were selected inside the training data, and the same test rows were used for the global, hard, soft and oracle variants.

A second strength is the direct ablation of global and segment-aware components. Performance was assessed using R^2, RMSE and MAE together with uncertainty intervals, paired tests and a seed-level bootstrap rather than a single point estimate.

Duplicate auditing, zero-overlap group-aware splits, router discrimination and calibration measures, and subgroup MAE provide additional checks on why the aggregate results changed. These analyses do not replace external validation, but they make the limits of the reported comparison clearer.

4.4 Limitations of the Study

The dataset did not include market-level variables such as exchange rates, tax policies, regional demand or macroeconomic conditions, although these factors can affect used car prices.

The dataset also lacked vehicle-specific information such as accident history, replaced parts, maintenance records, location, listing date, equipment package, color and overall condition. The absence of these variables limits the predictive information available to the models.

Only two- and three-segment structures were evaluated. Other numbers of segments, clustering algorithms or hybrid segmentation methods may produce different results. In addition, routing accuracy directly affects the performance of the segment-aware system.

The IQR-restricted analysis was conditional on a price range learned from the training target. Although the test distribution was not used to define the IQR limits, the true test price was still required to determine whether a test record belonged to the restricted cohort. Therefore, these results should not be interpreted as full-market deployment performance. The separate full-market evaluation was reported without target-based test exclusion.

CC: BY-NC-ND

Exact duplicates were removed before splitting, but predictor-identical vehicles with different recorded prices were retained because they may represent genuine unobserved variation. The group-aware check reduced the risk that repeated brand-model-age families alone created the result; however, it cannot rule out all dependence among listings or establish performance in a new market or time period. In addition, the study has not been validated on an independent external dataset or on a chronologically separated test cohort. Consequently, the repeated-split results establish internal robustness within the present dataset but do not establish generalizability across markets or stability under temporal distribution shift.

Subgroup findings were descriptive and based on one seed. Error varied by price, age, mileage, brand, fuel and transmission, and some categories contained few records. Aggregate performance should therefore not be interpreted as equal accuracy for every vehicle group.

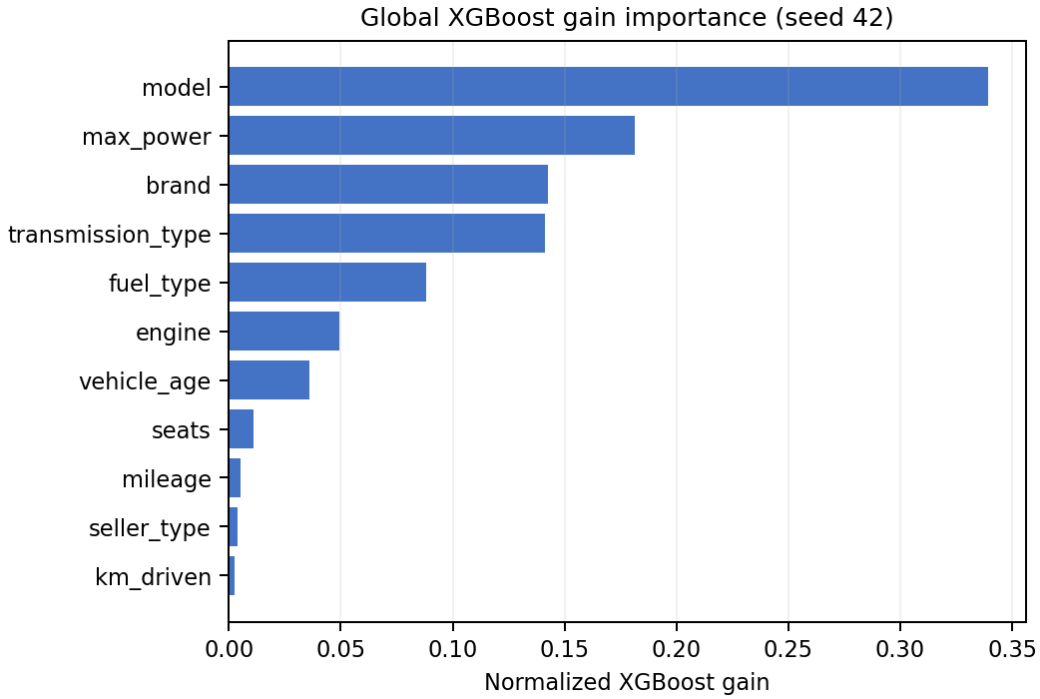

Figure 7. *Seed-42 global XGBoost feature importance based on normalized gain aggregated from one-hot columns to their source variables.*

Figure 7 reports normalized XGBoost gain, not permutation importance or a causal effect. The gain values were summed across one-hot levels belonging to the same source variable and normalized to sum to one before the leading variables were displayed. Model identity had the largest aggregated gain, followed by maximum power, brand and transmission type. Correlated predictors can nevertheless share or substitute for importance.

4.5 Future Work

Future studies may evaluate the framework on larger and more recent datasets. Because used car prices vary across countries, currencies and market conditions, external evaluation in different markets would provide stronger evidence of generalizability.

Additional variables, including accident history, maintenance records, location, listing date, equipment level and vehicle condition, may improve predictive performance.

Future work may also examine different numbers of segments and alternative clustering approaches, such as Gaussian Mixture Models, DBSCAN or hierarchical clustering. Segments could also be defined from vehicle features rather than only from price.

CC: BY-NC-ND

The routing stage may be improved through preregistered parameter tuning, alternative classifiers or calibration methods evaluated entirely inside training folds. Boundary-aware or uncertainty-aware combinations of expert predictions may also be examined.

TabNet, FT-Transformer and SSL-AutoEncoder require a separate fair benchmark with matched compute, common search spaces, repeated seeds and the same nested protocol before any comparative conclusion is drawn.

Future evaluation should add temporal and external-market test sets and report prediction intervals as well as point errors. Larger samples of rare brands and fuel types would also support more reliable subgroup assessment.

5. Conclusion

This study compared global and segment-aware regression models for used car price prediction using the CarDekho dataset. The original manuscript structure and model rationale were retained, while the revised analysis added exact de-duplication, sealed repeated splits, training-only topology selection, Duan retransformation correction and direct hard, soft and oracle ablations.

In the primary full-market evaluation, global XGBoost achieved mean R^2=0.8523, RMSE=349,460 INR and MAE=104,138 INR across ten splits. Soft gating achieved R^2=0.8733, RMSE=321,661 INR and MAE=99,196 INR. Paired Wilcoxon tests supported the average differences, although the variation in full-market RMSE remained large because rare expensive vehicles were distributed differently across splits.

Hard routing improved R^2 and RMSE but produced only a small mean MAE change relative to the global model. Soft gating reduced MAE further, and oracle routing showed additional headroom. In seed 42, paired bootstrap intervals included zero for R^2 and RMSE but not for MAE, so the evidence is strongest for a modest reduction in typical absolute error.

The secondary IQR-restricted results were more stable but showed almost no R^2 or RMSE difference between global and soft-gated models. The direction of the full-market improvement was retained in five zero-overlap brand-model-age group-aware splits. Router and subgroup diagnostics nevertheless showed that performance varied across segments and vehicle groups.

Global XGBoost therefore remains a strong and simpler practical baseline. Segment-aware soft gating can provide an empirical improvement when its added training, routing and monitoring burden is acceptable, but it is not uniformly superior and is not presented as a new algorithm. External and temporal validation are required before deployment claims can be made. The high variability of full-market RMSE across splits and the absence of external or temporal validation should therefore remain central considerations when interpreting these findings.

Data Availability Statement

The dataset analyzed in this study is publicly available through the CarDekho Used Car Dataset on Kaggle, as cited in the References section. Supplementary Material S1 contains the complete reviewer_revision.py analysis script, pinned requirements, per-split results, statistical tests, router and subgroup outputs, and the legacy implementation for provenance. The Kaggle data are not redistributed in the supplement.

Appendix A. Reproducibility Details

Table A1 consolidates the settings needed to reproduce the revised analysis. Parameters not listed for a library estimator retained that library version's default value. The delivery script and machine-readable outputs in Supplementary Material S1 remain the authoritative executable record.

CC: BY-NC-ND

Table A1. Reproducibility settings used in the revised analysis.

Component	**Setting**
Outer evaluation	Ten sealed 80/20 random splits; seeds 11, 23, 37, 42, 59, 71, 83, 97, 109 and 131. The outer test set was not used for preprocessing, topology, probability-method or parameter decisions.
Preprocessing	model levels with fewer than 25 observations in the current training fold mapped to Other; StandardScaler for six numerical predictors; OneHotEncoder(handle_unknown='ignore') for five categorical predictors. CatBoost used the same one-hot design matrix; native categorical handling was not used.
Segment candidates	2-Segment Quantile: 40th percentile; 3-Segment Quantile: 33rd and 67th percentiles; 3-Segment KMeans: three clusters on log1p(training price), n_init=20. Candidate chosen by three-fold training-only CV R^2.
Global/high expert XGBoost	n_estimators=250; learning_rate=0.05; max_depth=5; subsample=0.85; colsample_bytree=0.85; objective='reg:squarederror'; random_state=outer seed; n_jobs=1.
Low/middle expert CatBoost	iterations=250; learning_rate=0.05; depth=6; loss_function='RMSE'; random_seed=outer seed; verbose=0; thread_count=1; allow_writing_files=False.
XGBoost router	n_estimators=250; learning_rate=0.05; max_depth=5; subsample=0.90; colsample_bytree=0.90; eval_metric='mlogloss'; random_state=outer seed; n_jobs=1. Segment classes were balanced by down-sampling to the smallest training segment.
Probability selection	Raw XGBoost probabilities versus three-fold sigmoid calibration, selected by log loss on an internal stratified 20% subset of the outer-training data, then refitted on all outer-training records.
Target and retransformation	All regressors fitted log1p(selling_price). A three-fold out-of-fold Duan factor was estimated from training residuals and applied multiplicatively after expm1 inversion.
Legacy linear/tree screen	LinearRegression defaults; Ridge alpha=1.0; Lasso alpha=0.01 and max_iter=10,000; Random Forest: 150 trees, max_depth=18, min_samples_split=5, min_samples_leaf=2; Gradient Boosting: 200 trees, learning_rate=0.05, max_depth=4, subsample=0.85.
Legacy deep-model profiles	TabNet: n_d=n_a=16, n_steps=3, gamma=1.3, n_independent=n_shared=2, learning_rate=10^{-4}, at most 30 epochs, patience=6, batch=256. FT-Transformer: token=32, heads=4, blocks=3, feed-forward multiplier=2, dropout=0.15, AdamW learning_rate=5×10^{-4}, weight_decay=10^{-4}, at most 30 epochs, patience=6, batch=256. SSL autoencoder: hidden=128, latent=64, dropout=0.15, AdamW learning_rate=5×10^{-4}, weight_decay=10^{-4}, 20 pretraining and 30 fine-tuning epochs, patience=6, batch=256. These profiles are reported for provenance and are excluded from revised inference because the tuning/validation budgets were not equivalent.

CC: BY-NC-ND

Component	Setting
Software and hardware	Python 3.12.13; NumPy 2.3.5; pandas 2.2.3; SciPy 1.17.0; scikit-learn 1.6.1; XGBoost 3.0.5; CatBoost 1.2.8; CPU-only execution on an AMD EPYC 9V74 host with 9 logical CPUs and 15.64 GiB RAM.
Mean time per outer split	Topology selection 9.72 s; global fit 0.88 s; final segment experts and router 6.64 s. Times are hardware- and library-dependent.

Peer-review: External, Independent.

Acknowledgements:
The authors thank the anonymous reviewers and the Editorial Office for their constructive comments during the review process.

Declarations:

1. Statement of Originality:
This work is original.

2. Author Contributions:
Concept: RNK,HO; ***Conceptualization:*** RNK,HO; ***Literature Search:*** RNK,HO; ***Data Collection:*** RNK,HO; ***Data Processing:*** RNK,HO; ***Analysis:*** RNK,HO; ***Writing – original draft:*** RNK,HO; ***Writing – review & editing:*** RNK,HO.

3. Ethics approval:
Not applicable. This study used a publicly available secondary dataset and did not involve human participants, personally identifiable data, animals, or clinical interventions.

4. Funding/Support:
This work has not received any funding or support.

5. Competing Interests:
The authors declare no competing interests.

6. GenAI Usage Statement:
Generative AI tools were used to assist with language editing, document organization, and the preparation of reviewer-requested analyses. The authors reviewed the generated material and remain responsible for the study design, analysis, interpretation, and final manuscript.

7. Sustainable Development Goals:

☐ ☐ ☐ ☐ ☐ ☐ ☐

☐ ☒ ☐ ☐ ☒ ☐ ☐

☐ ☐ ☐ ☐

REFERENCES

[1] Pudaruth, S. (2014). Predicting the price of used cars using machine learning techniques. International Journal of Information & Computation Technology, 4(7), 753–764.

[2] Monburinon, N., Chertchom, P., Kaewkiriya, T., Rungpheung, S., Buya, S., & Boonpou, P. (2018).

CC: BY-NC-ND

Prediction of prices for used car by using regression models. 2018 5th International Conference on Business and Industrial Research (ICBIR), 115–119. https://doi.org/10.1109/ICBIR.2018.8391177

[3] Gegic, E., Isakovic, B., Keco, D., Masetic, Z., & Kevric, J. (2019). Car price prediction using machine learning techniques. TEM Journal, 8(1), 113–118. https://doi.org/10.18421/TEM81-16

[4] Cui, B., Ye, Z., Zhao, H., Renqing, Z., Meng, L., & Yang, Y. (2022). Used car price prediction based on the iterative framework of XGBoost+LightGBM. Electronics, 11(18), 2932. https://doi.org/10.3390/electronics11182932

[5] Huang, J., Yu, Z., Ning, Z., & Hu, D. (2022). Used car price prediction analysis based on machine learning. In Proceedings of the 2022 International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID 2022) (pp. 356–364). Atlantis Press. https://doi.org/10.2991/978-94-6463-010-7_37

[6] Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324

[7] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

[8] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

[9] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. Advances in Neural Information Processing Systems, 31, 6638–6648.

[10] Arik, S. Ö., & Pfister, T. (2021). TabNet: Attentive interpretable tabular learning. Proceedings of the AAAI Conference on Artificial Intelligence, 35(8), 6679–6687. https://doi.org/10.1609/aaai.v35i8.16826

[11] Gorishniy, Y., Rubachev, I., Khrulkov, V., & Babenko, A. (2021). Revisiting deep learning models for tabular data. Advances in Neural Information Processing Systems, 34, 18932–18943.

[12] Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? Advances in Neural Information Processing Systems, 35, 507–520.

[13] Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. Neural Computation, 3(1), 79–87. https://doi.org/10.1162/neco.1991.3.1.79

[14] Jordan, M. I., & Jacobs, R. A. (1994). Hierarchical mixtures of experts and the EM algorithm. Neural Computation, 6(2), 181–214. https://doi.org/10.1162/neco.1994.6.2.181

[15] Lloyd, S. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

[16] Surjandari, I., Dzikri, A., Dhini, A., Laoh, E., Pangesty, K. D., Aurora, P. S., & Ferrouzi, D. (2024). Stacked generalization with sequential-model based optimization for estimating used car valuation in Indonesia. Engineering, Technology & Applied Science Research, 14(5), 17239–17247. https://doi.org/10.48084/etasr.8226

[17] Madhusudhanan, K., Behrens, G., Stubbemann, M., & Schmidt-Thieme, L. (2024). ProbSAINT: Probabilistic tabular regression for used car pricing. 2024 IEEE International Conference on Big Data (BigData), 2179–2187. https://doi.org/10.1109/BigData62323.2024.10825740

[18] Bergmann, S., & Feuerriegel, S. (2025). Machine learning for predicting used car resale prices

CC: BY-NC-ND

using granular vehicle equipment information. Expert Systems with Applications, 263, 125640. https://doi.org/10.1016/j.eswa.2024.125640

[19] Fayyaz, I., Ali, G. G. M. N., & Khairunnesa, S. S. (2025). Advanced feature engineering and machine learning techniques for high accurate price prediction of heterogeneous pre-own cars. Vehicles, 7(3), 94. https://doi.org/10.3390/vehicles7030094

[20] Chen, H., Wang, H., Li, J., Yang, T., Wang, H., & Zhang, F. (2026). A research on the residual value of secondhand new energy vehicles based on CatBoost-BiLSTM model. Journal of Big Data, 13, 41. https://doi.org/10.1186/s40537-026-01374-z

[21] Netayawijit, P., Chansanam, W., & Sorn-In, K. (2026). An optimized data partitioning framework for benchmarking ensemble and linear regression models in used-vehicle price prediction. Engineering, Technology & Applied Science Research, 16(2), 32800–32809. https://doi.org/10.48084/etasr.16878

[22] Kaggle. (n.d.). CarDekho Used Car Dataset. https://www.kaggle.com/datasets/manishkr1754/cardekho-used-car-data (accessed July 20, 2026).

[23] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

[24] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

[25] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[26] Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. Proceedings of the 25th International Conference on Machine Learning, 1096–1103. https://doi.org/10.1145/1390156.1390294

[27] Duan, N. (1983). Smearing estimate: A nonparametric retransformation method. Journal of the American Statistical Association, 78(383), 605–610. https://doi.org/10.1080/01621459.1983.10478017.

CC: BY-NC-ND