

Trustworthy Residual Vehicle Value Prediction for Auto Finance

Mihye Kim¹, Jimyung Choi¹, Jaehyun Kim¹, Wooyoung Kim¹,
Yeonung Baek¹, Gisuk Bang¹, Kwangwoon Son¹, Yeonman Ryou¹, Kee-Eung Kim²

¹ Hyundai Capital Services, Korea

² Kim Jaechul Graduate School of AI, KAIST, Korea

{mihye.kim, jimyung.choi, jaehyun2.kim, wooyoung.kim}@hcs.com, keeung.kim@kaist.edu

Abstract

The residual value (RV) of a vehicle refers to its estimated worth at some point in the future. It is a core component in every auto finance product, used to determine the credit lines and the leasing rates. As such, an accurate prediction of RV is critical for the auto finance industry, since it can pose a risk of revenue loss by over-prediction or make the financial product incompetent through under-prediction. Although there are a number of prior studies on training machine learning models on a large amount of used car sales data, we had to cope with real-world operational requirements such as compliance with regulations (i.e. monotonicity of output with respect to a subset of features) and generalization to unseen input (i.e. new and rare car models). In this paper, we describe how we addressed these practical challenges and created value for our business at Hyundai Capital Services, the top auto financial service provider in Korea.

Introduction

In auto finance, a vehicle is the main object or purpose for which the financial service is provided. As such, the estimation of the residual value (RV) of the vehicle is a critical factor in determining the credit lines and the leasing rates. As an example, for an auto loan, the estimated market price of the secured vehicle is used to determine the credit line. As another example, for an auto lease, the predicted future value of the vehicle at the time of its return is used to determine the monthly rate. If the RV of the leased vehicle is over-predicted than the actual value at the time of return, it would impose a risk of revenue loss, while an under-prediction will push up the lease rate and thereby potentially lose the customer to competitors. Thus, an accurate prediction of RV is one of the most important business requirements of every auto financial service.

Accurate RV prediction is also critical for the used car market as a whole, beyond the auto finance business. In most countries including Korea, the used car market is commonly recognized as a *lemon market*, where information asymmetry exists between sellers and buyers. Buyers lack sufficient information for determining a fair price for the vehicle and thus worry about paying more than what it is worth. This leads to buyers making their best attempt to pay less, and

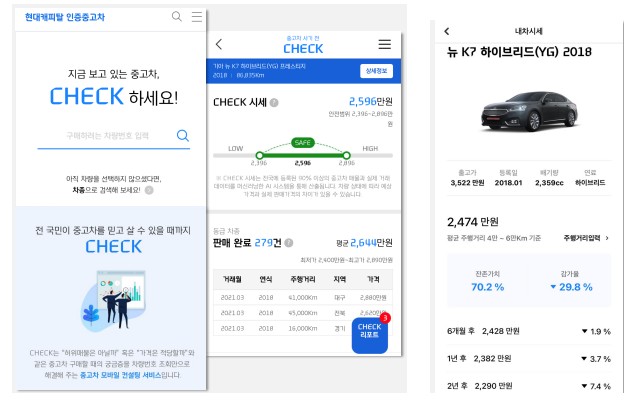


Figure 1: Screenshots of the used car price estimation service menu in the HCS mobile app, for potential buyers (left) and potential sellers (right). The price estimation can be made at the individual car level by supplying the license plate number.

auto dealers consequently sell cars of inferior quality to match the price. Ultimately, both the sellers and the buyers become dissatisfied with the used car market, resulting in a decreasing number of market participants. This unfortunate market behavior is difficult to rectify unless the information balance between sellers and buyers is restored. This is also one of the goals behind our work, to make accurate predictions of used car prices available to ordinary consumers to protect them against exploitative dealers, thereby contributing to a healthier and more robust market for used cars. This ultimate goal is also important for us at Hyundai Capital Services (HCS), the top auto financial service provider in Korea ever since its launch of the retail (installment) finance business in 1996.

As a part of our effort, we have been providing a personalized used car price estimation service through a mobile channel since 2018, shown in Figure 1, to improve customer engagement. This service leverages a deep neural network (DNN) model trained on proprietary used car sales data. At the time of the service launch, it was the first to adopt deep learning and thus attracted a lot of attention. Yet, the model was unsuited for the core business, since it fell short in many operational requirements: as shown in the left pane of Fig-

ure 7, although the model captures the general trend of value depreciation with respect to mileage and age of a specific car model, the monotonicity is clearly violated in many local regions, as can be seen from the spikes and valleys. In addition, the model was not able to handle novel or rare car models. This paper describes how we addressed these practical limitations in the baseline model and created business values by integrating the improved model into the core component of auto finance services, handling tens of thousands of monthly cases and increasing quarterly profit by several million US dollars since Q3 2022.

Related Work

The vehicle RV prediction can be naturally formulated as a regression task in supervised learning, using many sources of used car sales transaction data readily available. There is thus a vast amount of literature on the topic, ranging from classical linear regression models to deep neural networks, and we only highlight some of the most relevant work for our approach.

Most notably, Lessmann and Voß (2017) report a systematic study on empirical performances of 19 regression models on a proprietary dataset consisting of 4.5 million instances covering 6 car models from the same manufacturer. The regression models used for experiments range from classical linear regression models, non-linear models, and ensemble models. The study finds that non-linear models and ensemble models perform much better than linear models, which is a natural result given that the value depreciation is inherently non-linear. This finding is consistent with other studies on RV prediction, e.g. support vector machine regression (Lessmann, Listiani, and Voß 2010) and neuro-fuzzy systems (Lian, Zhao, and Cheng 2003; Wu, Hsu, and Chen 2009). More recent work adopts an advanced training paradigm to learn a good representation of input for improving RV prediction accuracy (Rashed et al. 2019).

However, non-linear models often suffer from explainability. In highly regulated domains such as finance, the model needs to be fully audited before its deployment. Due to this operational requirement, linear regression models are still popular in practice. The AutoCycle™ by Moody's analytics uses logistic regression and quantile regression models to provide a commercial service for vehicle value prediction (Hughes et al. 2016). It allows forecasting of car prices under various macroeconomic scenarios such as changes in the unemployment rate and gas prices. The model was reported to achieve an R-squared of 0.89 on 6.4 million test instances. Yet, being a linear prediction model by its nature, it relied on heavily engineered features that require substantial expert knowledge, despite a huge amount of data (approximately 31 million instances from the National Automobile Dealers Association) being available for training and validation.

Our work focuses on training an expressive non-linear model for RV prediction from a large amount of data with the following desiderata: (1) enforce operational requirements of the model by introducing monotonicity constraints (e.g. RV prediction should be monotonically decreasing with mileage and age) for fairness and transparency, and (2) make

the model generalize well on out-of-distribution data (e.g. new car models released to the market by the manufacturers). To the best of our knowledge, these two requirements are yet to be simultaneously addressed for a commercial-grade machine learning model, at least for the RV prediction task of vehicles in auto finance.

Task Description

Our objective was to develop an expressive (i.e. non-linear) machine learning model that accurately estimates vehicles' current and future market prices while addressing the following operational requirements:

1. For the same car model, the RV ratio (i.e. the price-to-MSRP ratio) should be lower if the mileage (total distance traveled) is higher and if the age is older. Also, the RV ratio should be higher for more recently released models and for lower-level trims. The financial authorities regulate whether the finance companies' loan limits or pricing standards are adequate and well-explained to protect financial consumers from undue damages. As a vehicle price prediction model establishes standards for setting the financial product's credit lines and pricing in the auto finance industry, the model used needs to be highly accurate as well as have a clear rationale for its logic.
2. The model should be able to estimate RVs for all car models including those new on the market with no history of used car trading. The actual used car sales data is the main source of information for training the model, but new car models are continuously released quarterly or semi-annually, and these newer models inevitably lack sufficient sales data for reliable training. Financial companies, nonetheless, face the challenge of having to estimate the market price of fresh models even before there is any data available to add to the training set, since the lease product requires predicted RV at the end of the lease in order to determine monthly payments.
3. The model needs to be updated continuously and regularly. Since the economic situation is always changing, more recent used car sales data is more accurate reflection of how much the car is worth today, compared to older ones. It is important to maintain the prediction accuracy by regularly obtaining the latest data and updating the model.

The model is trained to predict the RV ratio, i.e. the ratio of the used car price to the MSRP of the new car. As for the training dataset, we secured about 1.8 million used car sales transaction records from 600 vehicle models which were actually traded in Korea over the past 12 years. Each transaction record contained a number of attributes such as vehicle specifications (manufacturer, brand, model, new car price, etc.), vehicle usage information (mileage, age, whether it was for rental, etc.), and market supply & demand (no. of used cars available on the market for sale, new car sales volume, used car searches on platform, time taken for sale, etc.). There were a total of 71 candidate features, composed of 40 vehicle specifications, 26 vehicle usage information,

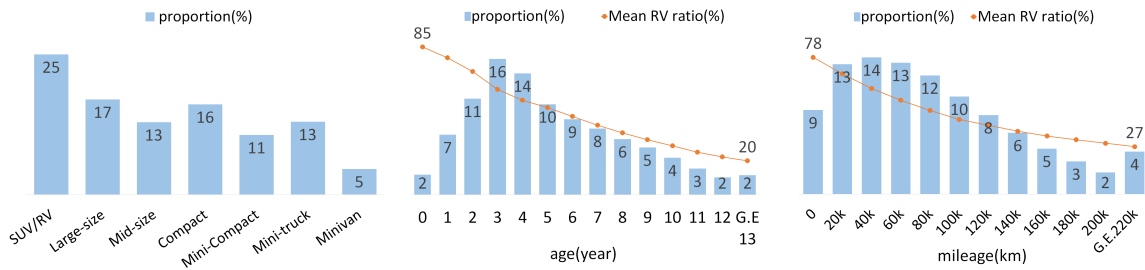


Figure 2: Statistics chart of training dataset

and 5 market condition information. Additional information on the summary statistics of the training data is provided in Figure 2

Methodology

Dataset

We used a proprietary dataset of used car sales transactions, published monthly by the dealers’ association in Korea. We used the actual sale prices as the target of the model for training. We filtered out instances that were obviously incorrect or beyond the scope of our interest. For example, transactions of commercial vehicles such as taxis and trucks, RV ratio greater than 1.0, and yearly mileage greater than 200k km (124k miles) were removed, which comprised 6% of the original data.

Anomaly Detection

Removing anomalies is a crucial process in exploratory data analysis (EDA) for improving a model’s prediction accuracy. It was even more important in our case because most vehicle sales price data were entered manually by dealers. This introduces opportunities for human errors and even fraudulent cases such as forged sales price to avoid taxes. Thus it was essential to filter out abnormal instances in the dataset before training the RV prediction model. Although there are many promising anomaly detection methods, we limited ourselves to comparing Extended Isolation Forest (EIF) and Local Outlier Factor (LOF) which do not rely on any distributional assumption.

The EIF method is a tree-based algorithm effective for multi-dimensional data with large volumes. It calculates the leaf node (terminal node) distance for each instance to become isolated in order to derive the score. The score is higher if the distance (depth) is shorter, and the outliers are defined based on the score value (Hariri, Kind, and Brunner 2019). The LOF method defines anomaly points by taking into consideration the relative density among the clusters. It calculates the local density based on k-nearest neighbors (KNN), and compares the local density of one object against the local density values of its neighbors, categorizing those with large gaps as outliers (Breunig et al. 2000).

We measured the precision and recall on a subset of data, restricted to one particular vehicle model with the largest sales volume in the new and used car market. For this subset of data, we recruited domain experts to label the instances

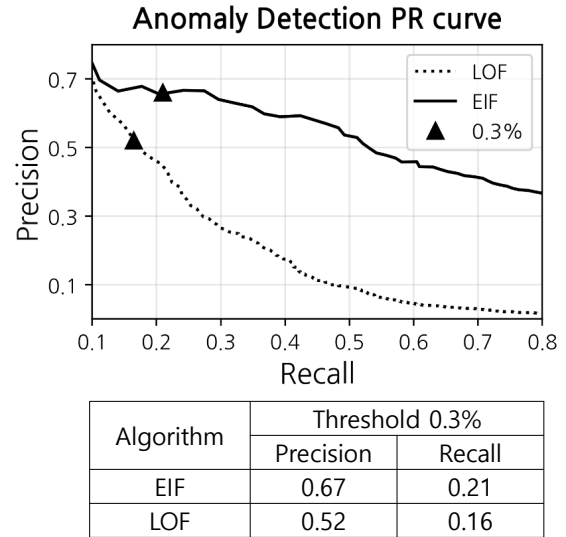


Figure 3: Performance comparison of anomaly detection algorithms.

as normal vs. abnormal to compute the performance metrics. This held-out labeled dataset contained approximately 60k instances. When comparing the performances of the two methods at the same threshold (0.3%), EIF performed better than LOF (Figure 3).

In addition, the outlier detection performance of the two algorithms is compared to a more traditional method, IQR, which is a method of setting the range for each major feature (age, mileage) using the interquartile range, and then defining the object outside the range as an outlier (Vinutha, Poornima, and Sagar 2018). EIF outperformed IQR in anomaly detection in individual vehicle models while LOF showed a tendency to classify some of the normal data as outliers (Figure 4). Therefore, we adopted EIF for carrying out multi-dimensional anomaly detection.

Vehicle Model Code Vectorization leveraging DNN

There are about 500 vehicle models that need to be covered by our RV prediction model, some of which have sufficient data while others do not. It was one of the important requirements for us to predict the RV as accurately as possible for all vehicle models, including the ones that lacked sufficient

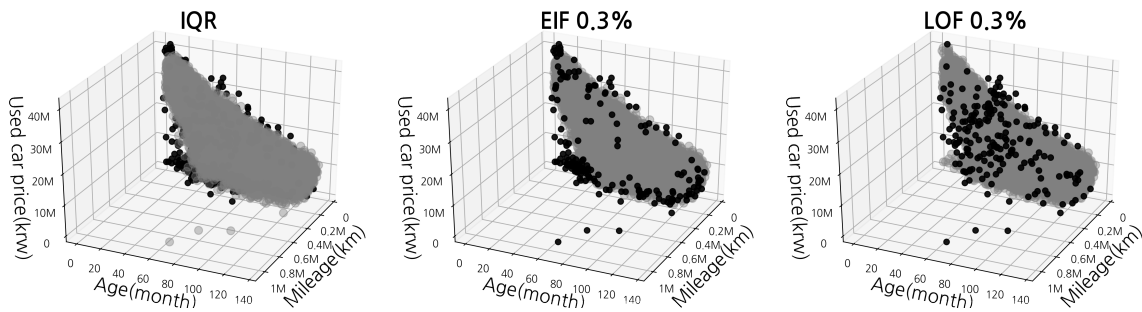


Figure 4: Comparison of anomaly detection algorithms.(Black Dots: Outliers classified by each method.)

amount of training data. For this purpose, it was required to vectorize the ‘Vehicle Model Code’, a categorical variable that distinguishes individual vehicle models, and use it on the prediction model. Even though the generally used one-hot encoding converts the categorical values to numerical values, it had limitations in reflecting the unique characteristics of each vehicle model.

Therefore, we employed the embedding method. Embedding has the advantages of generating denser representations compared to one-hot encoding, and showing not only characteristics of each vehicle model but also similarities between vehicle models with finite-dimensional vectors.

The embedding was obtained by plugging an embedding layer into a neural network, trained to predict RV ratio (Figure 5). Embedding dimension was set to 2 for the sake of interpretability and visualization for domain experts. We considered various architectures for the neural network (e.g. number of layers and hidden units), and the first batch of candidates was selected solely based on the RV ratio prediction accuracy of the neural network. They were then reviewed by domain experts, and further narrowed down by visually inspecting whether the overall patterns of embedding was consistent with their knowledge. The final embedding was selected by the RV ratio prediction accuracy of the CatBoost model.

Using only 2 dimensions for embedding, the performance of the RV prediction model was significantly improved. When the embedding vectors of vehicles were visualized, they could be clustered in a meaningful way (Figure 6). We used this clustering result to assign embedding vectors to vehicles that could not be projected onto embedding vector space due to a lack of data. For example, if the specification information of a newly released vehicle is similar to that of vehicles in a specific cluster, the centroid of the corresponding cluster is assigned as an embedded vector value of the new vehicle. For a car model with complex characteristics that cannot be determined by a single cluster, we used two or more selected clusters. Using the centroids of each cluster, we found a new point that reflects the cluster information and use it as an embedding vector. By doing so, we were able to solve the cold-start problem caused by the lack of data on newly launched and rarely traded car models.

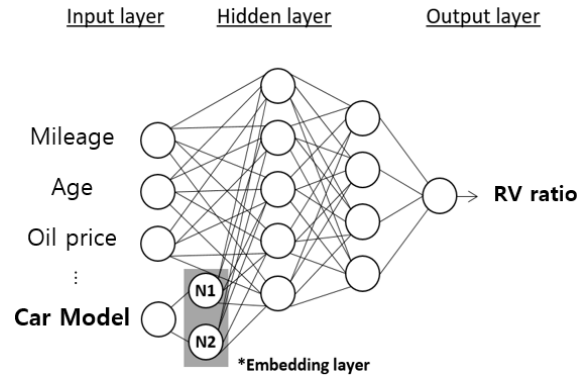


Figure 5: Neural network architecture for obtaining vehicle model embedding.

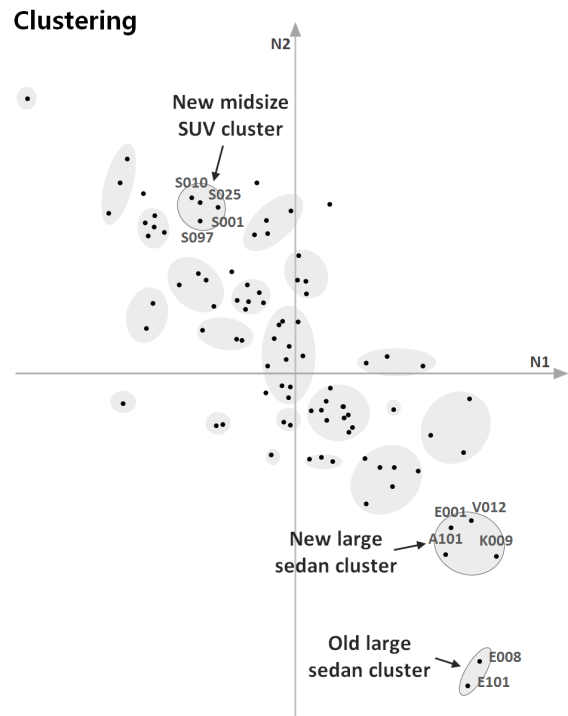


Figure 6: Visualization of vehicle embedding vectors. Vehicle models are anonymized.

Training Boosted Tree Models

We selected tree-based models for the following reasons: First, we aimed to improve the prediction accuracy for vehicle data by capturing complex complicated non-linear relationships between variables. In addition, the model should be able to conveniently reflect the monotonicity of the target variable with respect to a subset of features. We trained the models using XGBoost, one of the most powerful and popular tree-based machine learning methods, and CatBoost, which specializes in learning categorical variables. In particular, CatBoost has the advantage of mitigating data leakage and recognizing categories with similar characteristics in groups because it processes categorical features via the mean target encoding method (Dorogush, Ershov, and Gulin 2017). This is particularly helpful for the vehicle data since they contain various categorical features such as vehicle model code, color, and maker. We compare the performances of XGBoost and CatBoost in this section.

We trained the model by setting the RV ratio as the target variable. From a total of 71 features, we determined that 38 of them being of low relevance or replaceable by other derivative features. Another 20 of them had too many missing value cases. We thus selected 13 features from the dataset, and augmented them with 8 additional features from side information (e.g. how long since the car model generation was first released in the market), resulting in a total of 21 features. The selected features were 5 vehicle specifications, 11 vehicle usage information, and 5 market demand/supply information.

As mentioned in the Task Description section, we set the constraints checking the monotonicity when creating a tree, in order to satisfy the monotonic relationship between a subset of features and the RV ratio. Monotonic constraints were given to 4 features; age (month), mileage, new car price, and model year. In the model that did not account for monotonicity, the price prediction shows significant fluctuation depending on the age and mileage, as can be seen in the left graph of Figure 7. On the other hand, in the model we developed with monotonicity constraint, price prediction shows monotonicity for specific variables, as displayed in the graph on the right. Monotonicity enforces the model to accord with common sense; the more high-end options the vehicle has, and the more aged the vehicle is, the faster its depreciation. Also, the encoded vectors of vehicle model code derived by leveraging the DNN model were used as input features for training the boosted tree models. The final model consists of approximately 4500 trees, with the maximum depth of 9.

Results and Discussion

We conducted a performance evaluation of the vehicle price model for out-of-time (OOT) data sets, and used the Mean Absolute Error (MAE) of the actual RV ratio and model-predicted RV ratio as evaluation metrics. As we know, MAE is defined as $\sum_{i=1}^n |y_i - x_i|/n$ where y_i is the predicted value, x_i is the observed value and n is the total number of data points. Also, by defining $MAE_{car-seg}$ as below, the offset effect for each car model was aggregated in calculating MAE:

Model	MAE(%)		Coverage	Monotonicity
	Total	Car-Seg		
Baseline DNN	6.87	6.55	77%	Unsatisfied
DNN + XGBoost	3.82	1.62	100%	Satisfied
DNN + CatBoost	3.64	1.32	100%	Satisfied

Table 1: Comparison of vehicle price prediction accuracy between the baseline DNN model, and the boosted tree models (XGBoost and CatBoost).

Car Segments	MAE(%)	
	DNN+XGB	DNN+CAT
SUV/RV	3.24	3.15
Large-size	3.01	2.92
Mid-size	3.65	3.55
Compact	4.17	4.03
Mini-compact	4.24	4.01
Mini-truck(commercial)	4.91	4.67
Minivan	5.96	4.98

Table 2: Performance by vehicle model of the prediction models leveraging XGBoost (XGB) and CatBoost (CAT).

$$MAE_{car-seg} = \frac{\sum_i |\overline{y_{car_i}} - \overline{x_{car_i}}|}{n_{car}} \quad (1)$$

where $\overline{y_{car_i}}$ is the average predicted value of car model i , $\overline{x_{car_i}}$ is a mean observed value of car model i , and n_{car} is the total number of car models in the data set. The performance of models that were developed using each algorithm is shown in Table 1.

Compared to the baseline model (DNN only), currently deployed for used car estimation service in Figure 1, the new models showed a substantial improvement in error rate while satisfying the monotonicity for key variables complying with regulations; a 44% error reduction compared to the baseline model by XGBoost, and a 47% error reduction by CatBoost. There were even more improvements in the error rate by car segment than the total errors of individual vehicles (a 76% error reduction by XGBoost, an 80% error reduction by CatBoost). It is due to the fact that the characteristics of the vehicle model were reflected in the models through embedding vectors. It was also confirmed that CatBoost outperformed XGBoost (Table 2). This is a natural result, given that CatBoost handles categorical features better than XGBoost, and our training data contained many categorical features. Therefore, we selected DNN+CatBoost as the final model. In the final model, the overall errors decreased 47% and errors by vehicle model decreased 80% compared to the baseline model (DNN only) while satisfying the monotonicity of key features (Figure 7).

Lastly, to verify the effectiveness of removing anomalies, we used raw data to train the prediction model leveraging the DNN+CatBoost model. When compared to the model trained with raw data, the model trained with cleansed data preprocessed using anomaly detection decreased MAE by 0.87%, from 4.51% using the raw data to 3.64% using the cleansed data.

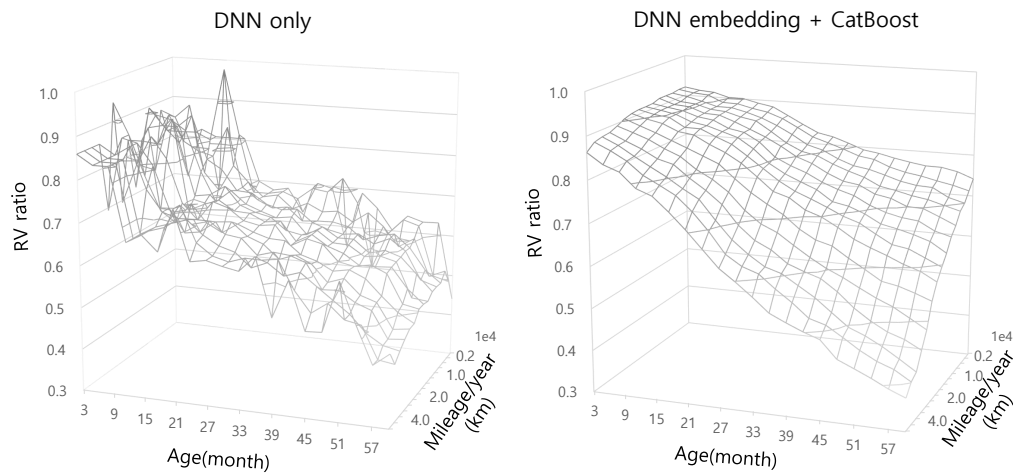


Figure 7: Comparison of predicted RV ratios projected on age and mileage per year.

In summary, we were able to develop a new model with significantly improved accuracy. The baseline model could not be applied for our company’s core business processes (calculating the lien value for used car loans, predicting vehicle return prices in order to set the monthly lease rates, and leveraging reserves) due to its issues in the prediction accuracy, the cold-start problem with new car models, and the lack of monotonicity guarantee for key features. It was therefore only used for the casual used car price estimation service available through our company’s mobile channel. The new model resolves all three issues above, improving prediction accuracy by at least 15% for all car models. The business unit thus approved our model for deployment in early 2022.

Application Use and Payoff

First, we used the model to determine the residual value of our leased vehicles. As mentioned at the beginning of this paper, an incorrect estimate of residual value can have a direct impact on our profit and loss. At the end of a lease, customers can choose to return their car or buy it out. If the residual value we have estimated is higher than the market value, it is more likely for the customer to return the car. When we end up selling the returned car, the difference between the predicted value and actual market value would be added to our loss. In order to reduce this loss, we could simply set the residual value at the end of the lease as low as possible, but this would have made monthly lease payments expensive, resulting in less consumers deciding to lease a car with our company. The more accurately we can predict the residual value of a returning car, the less the loss caused by a prediction error. By replacing the baseline model with the one that we developed, we were able to offset our loss caused by an error in the predicted RV by approximately 21%. The gap between the forecast value and the actual price at sale was reduced to half the original level. This enabled us to reduce our reserves by a significant amount (several millions of dollars in Q3 2022 alone).

Second, our model was used as the benchmark in deter-

mining the credit line on loans provided for buying a used car or those secured by a used car. By offering a credit line that approximates the value of the car, we can reduce the account’s probability of default, and also prevent fraud cases where someone may take out a loan for an amount greater than what the vehicle is actually worth and intentionally default on the loan. Another benefit is that that we can encourage customers to spend reasonably, within their ability to repay.

Third, we used the model to offer a service allowing customers to find out what their car is worth on our existing mobile application, “Hyundai Capital.” Car prices estimated from the previous, DNN-based model were replaced with the ones from the newly developed model. This enabled us to offer more accurate forecast values to mobile application users. By using this service when planning to sell their car, customers can protect themselves from being taken advantage of by shady dealers. The service also helps customers decide to return the car at the end of a lease. We expect our enhanced service will attract more customers to our app and build stronger customer trust.

Lastly, we are selling the vehicle price database generated from our RV prediction model to used car online market companies via partnership. We have entered into contracts where we provide the partner companies with prices predicted by our model in return for lowering our advertising costs by 10-20%. We are steadily expanding these partner companies as we further solidify our position as the market leader in price forecasting.

Other than the areas of our business, we also use the model to better protect consumers and improve transparency in the used car market. As mentioned in the introduction, the problem of information asymmetry in the used car market has often made used car buyers susceptible to unfair deals. In order to prevent this, the South Korean government has established a website hosting a standard platform for used car companies (www.car365.go.kr). Used car prices estimated by our company are one of the standard prices shared on the government website. Aside from sharing our prices through

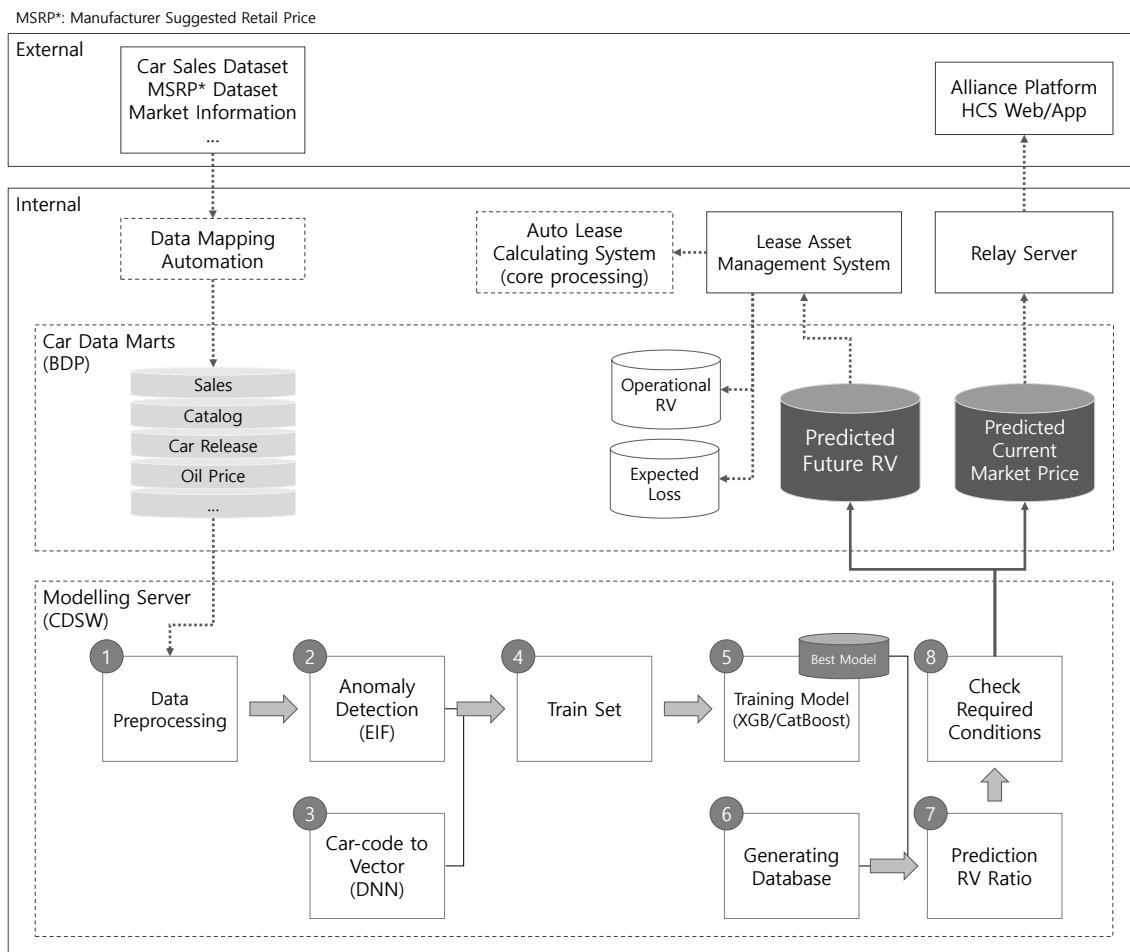


Figure 8: A workflow diagram illustrating automated operations of our vehicle price predicting model.

the government agency, we are also engaged in our own efforts to further assist consumers. A case in point is our service named “What to check before buying a used car” available on our Certified Pre-Owned car site. As its name suggests, the service helps consumers check before the purchase whether the car they have their eye on does actually exist, and whether the car is worth the price tag. We are enhancing the trustworthiness of our car price forecasts through the improvement we have made to the model, and thereby contribute to stronger protection of consumer rights.

In addition, in order to maintain a robust and healthy market for used cars, the Korean government requires validation when finance companies issue used car loans. The car prices predicted by our model have provided resources for setting the standard in this regard. With its enhanced accuracy, the newly launched RV model is expected to aid credit line-related risk management in the entire used car market.

Deployment and Maintenance

We completed the model development in 1Q, 2022. In 2Q, we deployed it into the HCS internal system and reviewed which business areas needed RV forecasting, and finally ap-

plied the model to core business processes in 3Q. Currently the new model handles several tens of thousands of customers each month for lease and used car loan products. A similar number of customers are also using the “Check out My Car’s Value” service on Hyundai Capital’s mobile channel.

In order to maintain our RV prediction model’s performance by keeping it up-to-date, we designed a workflow for retraining with the latest transaction data added every month. We have built the model’s operations architecture by setting up a vehicle data mart on our internal big-data platform and developing automated process modules (Figure 8). Vehicle data marts are composed mainly of the transaction data from the used car market and sales data of vehicles whose lease contracts have matured. Market information such as oil prices and vehicle shipments is added as well. These are used for retraining the model. The model development process described above was also modularized so that the model was automatically retrained in the modeling server. The model produces two outputs, the used cars’ current price estimates and their future residual value predictions. The current price data is updated monthly on the

Hyundai Capital price inquiry service platform through a relay server. Future price data is loaded through the internal strategic management system, and used for offering automotive lease products to customers.

Once the RV prediction model is retrained, it goes through a safety measure that automatically monitors the calculated RV ratios to review whether they are appropriate before delivering them to the relevant systems. The adequacy check process is as follows. There are two types of check points: model performance validation and operational stability check (health check). Model validation uses MAE as an evaluation index to check model performance and its level of variation from the previous month. In addition, the monotonicity is checked based on the items (age, mileage, trim) that are essential to satisfy the monotonicity of the residual price within the same vehicle model. The stability check confirms the level of change in the RV ratio by vehicle type compared to the previous month and the appropriateness of the vehicle depreciation. The change in the number of vehicles whose market price can be calculated is also checked against the total number of vehicle types traded in the market.

Conclusion

Predicting vehicle RV is a critical factor in the auto finance business, and for this reason we dedicated our efforts on enhancing the accuracy of the prediction model. While the single most important goal for our research was to improve the prediction performance of our model, in order to be able to apply it to a wider range of business areas, we also aimed to satisfy the monotonicity of key features as well as overcome the issue of data shortage for some models (“cold-start problem”). In order to address these challenges, we sought to upgrade model development across its entire stages, and enhanced the model by applying new techniques. We leveraged anomaly detection to sort out the labels, and also applied neural network in order to ensure embedding of the vehicle model. We also trained tree-based machine learning methods to achieve a significant jump in the model’s prediction accuracy (overall errors decreased 47%, errors by vehicle model decreased by 80%), at the same time as resolving the monotonicity and cold-start problems.

Through this research, we were able to enhance the reliability of our new model and thereby apply it to the core business of our company, achieving meaningful business results. As the top auto finance service provider in Korea, we are strongly committed to making reliable used car prices widely available in the market and thus protecting the right of the customer, and this research allowed us to contribute further in this respect. Going forward, the team at Hyundai Capital Service will continue to study and leverage the latest AI techniques to maintain and enhance the reliability of our residual value prediction.

References

Breunig, M. M.; Kriegel, H.-P.; Ng, R. T.; and Sander, J. 2000. LOF: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 93–104.

Dorogush, A. V.; Ershov, V.; and Gulin, A. 2017. CatBoost: gradient boosting with categorical features support. In *Workshop on ML Systems at NIPS*.

Hariri, S.; Kind, M. C.; and Brunner, R. J. 2019. Extended isolation forest. *IEEE Transactions on Knowledge and Data Engineering*, 33(4): 1479–1489.

Hughes, T.; Malone, S. W.; Brisson, M.; and Vogan, M. 2016. Introducing AutoCycleTM: Residual Risk Management and Lease Pricing at the VIN Level. Technical report, Moody’s Analytics.

Lessmann, S.; Listiani, M.; and Voß, S. 2010. Decision support in car leasing: a forecasting model for residual value estimation. In *International Conference on Information Systems*.

Lessmann, S.; and Voß, S. 2017. Car resale price forecasting: The impact of regression method, private information, and heterogeneity on forecast accuracy. *International Journal of Forecasting*, 33(4): 864–877.

Lian, C.; Zhao, D.; and Cheng, J. 2003. A fuzzy logic based evolutionary neural network for automotive residual value forecast. In *International Conference on Information Technology: Research and Education, 2003. Proceedings. ITRE2003.*, 545–548. IEEE.

Rashed, A.; Jawed, S.; Rehberg, J.; Grabocka, J.; Schmidt-Thieme, L.; and Hintsches, A. 2019. A Deep Multi-task Approach for Residual Value Forecasting. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 467–482. Springer.

Vinutha, H.; Poornima, B.; and Sagar, B. 2018. Detection of outliers using interquartile range technique from intrusion dataset. In *Information and decision sciences*, 511–518. Springer.

Wu, J.-D.; Hsu, C.-C.; and Chen, H.-C. 2009. An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference. *Expert Systems with Applications*, 36(4): 7809–7817.