Comparative Analysis of an Explainable Ensemble of Multi-View Deep Learning Models for Future Price Prediction of Pre-Owned Cars

Introduction:

The rapid expansion of the pre-owned vehicle market necessitates advanced price prediction tools capable of accurately and reliably assessing vehicle values. While traditional statistical methods have historically served this purpose, the advent of deep learning, particularly multi-view architectures, has ushered in new possibilities for enhanced prediction accuracy and the ability to leverage diverse data sources. However, the inherent lack of transparency in many deep learning models presents a significant challenge for stakeholders (including buyers, sellers, and financial institutions) who demand interpretable decision-making, especially in the high-stakes automotive market where pricing directly impacts financial outcomes.

This research introduces a novel ensemble approach that integrates multiple deep learning models, each designed to analyse distinct "views" or facets of vehicle data, while prioritizing explainability through built-in interpretation mechanisms. By synthesizing information from diverse data sources—spanning historical pricing trends and market dynamics to vehicle-specific attributes, visual condition, and textual descriptions—the ensemble architecture aims to capture nuanced patterns and interdependencies that single-view models might miss. Furthermore, the integration of explainability techniques seeks to transform these complex models from opaque predictors into transparent decision support systems, empowering stakeholders to comprehend and trust the logic underpinning price predictions. This approach addresses the critical need for both accuracy and transparency in automotive valuation.

The significance of this work lies both in its potential to increase prediction accuracy and its practical relevance to real-world applications where decision transparency is paramount. By bridging the gap between cutting-edge deep learning methodologies and interpretable predictions, this research contributes to both the theoretical advancement of multi-view ensemble models and their practical deployment within the automotive industry. The exponential growth of the used car market, driven by the demand for cost-effective transportation, underscores the need for accurate and reliable pricing strategies. Traditionally, machine learning models have predominantly relied on structured numerical data. In contrast, deep learning techniques offer the capacity to leverage multi-view inputs, including image and textual data, enabling a more holistic representation of a vehicle. However, the limited interpretability of deep learning models remains a key obstacle to their widespread adoption in scenarios requiring high-stakes decisions.

I. Traditional Machine Learning Approaches

Early attempts at price prediction employed regression models such as Linear Regression, Decision Trees, and Random Forest. These methods established a baseline for performance, providing relatively simple and interpretable models. However, they often struggled to fully account for the complex non-linear relationships and interactions among various predictive features in the automotive market. For instance, Linear Regression assumes a linear relationship between features and price, which may not hold true when considering factors like brand prestige or the non-linear impact of mileage on depreciation. Decision Trees and Random Forests can capture some non-linearities but may be limited in their ability to generalize to unseen data and capture subtle patterns present in high-dimensional datasets.

II. Deep Learning for Price Prediction

Recent breakthroughs in deep learning have led to the development of powerful architectures capable of automatically learning complex feature representations from data. Convolutional Neural Networks (CNNs) have shown exceptional promise in image-based vehicle condition assessment extracting features related to damage, wear, and overall appearance. Recurrent Neural Networks (RNNs) and Transformer networks have demonstrated their effectiveness in sequential data analysis, making them suitable for processing textual descriptions of vehicles and time-series data of price trends Hybrid models that combine information from multiple data modalities, such as integrating CNN-extracted image features with RNN-processed textual features, have emerged as a promising approach. However, the challenge of interpretability persists as a significant limitation in many deep learning models, which often function as "black boxes," making it difficult to understand the reasoning behind their predictions.

III. Multi-View Data Representation

To leverage the richness and diversity of automotive data, the proposed approach considers the following data views, each capturing a different aspect of a vehicle:

III.A Structured Data:
Encompasses numerical and categorical features such as mileage, vehicle age, brand, model, transmission type, body type, fuel type, engine capacity, and historical price trends. These data types are typically well-suited for tabular data processing methods.

III.B Textual Data:
Information extracted from seller descriptions, online reviews, and expert opinions using Natural Language Processing (NLP) techniques. This includes word embeddings that represent words as vectors and contextualized language models that capture the meaning of words in their surrounding context.

III.C Image Data:
Visual information obtained from vehicle images, which can be processed to evaluate vehicle condition and aesthetic attributes.

IV. Deep Learning Architectures

The proposed ensemble model integrates the following deep learning architectures, each tailored to a specific data view:

IV.A CNN for Image Analysis:
Convolutional Neural Networks are employed to automatically extract features from vehicle images, enabling the assessment of visual attributes such as dents, scratches, rust, paint condition, and overall

vehicle appearance. These features can provide valuable insights into the vehicle's condition and influence its perceived value.

IV.B RNN and Transformer for Text Analysis:

Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, and Transformer models are utilized to process textual descriptions, extracting relevant information such as vehicle features, condition descriptions, seller sentiment, and subjective opinions expressed in reviews. Attention mechanisms within Transformer models allow the model to weigh the importance of different words and phrases in the text.

IV.C Multi-Layer Perceptron (MLP) for Structured Data:

A Multi-Layer Perceptron, a type of feedforward neural network, is designed to process structured data, learning complex relationships between numerical and categorical features to predict the base price of the vehicle.

IV.D Ensemble Model:

The outputs from the CNN, RNN/Transformer, and MLP are combined using a weighted averaging mechanism to produce the final price prediction. The weights assigned to each model's output can be learned during training or determined based on expert knowledge to optimize the ensemble's performance.

V. Experimental Setup

Dataset:
The dataset comprises a combination of publicly available datasets (e.g., Mendeley, CarDekho, Kaggle) and potentially proprietary data from automotive marketplaces or dealerships. The dataset includes a diverse range of vehicle makes, models, years, and conditions, ensuring the model's generalizability.

Data Preprocessing:
Extensive data preprocessing is performed to ensure data quality and suitability for model training. This includes handling missing values, encoding categorical features (e.g., using one-hot encoding), scaling numerical features (e.g., using standardization or normalization), and addressing outliers.

Evaluation Metrics:
Model performance is rigorously evaluated using a range of metrics that capture different aspects of prediction accuracy.

Common metrics include:
Mean Absolute Error (MAE):
The average absolute difference between predicted and actual prices, providing a measure of the average magnitude of errors.

Root Mean Squared Error (RMSE):
The square root of the average of the squared differences between predicted and actual prices, giving more weight to larger errors.

R-squared (R^2):
A statistical measure of how well the model explains the variance in the target variable (price), indicating the goodness of fit.

V.D Baseline Models:
The proposed model's performance is compared against a set of baseline models to establish a performance benchmark and demonstrate the effectiveness of the multi-view ensemble approach.

Baseline models include:
Traditional machine learning models (Linear Regression, Decision Trees, Random Forest, Gradient Boosting)

Single deep learning architectures (CNN, RNN, LSTM, MLP)

VI. Conclusion

The Explainable Ensemble of Multi-View Deep Learning Models for Future Price Prediction of Pre-Owned Cars leverages the capabilities of deep learning to significantly improve both the accuracy and interpretability of price forecasts. By integrating diverse data types, including tabular data, time-series data, image data, and text data, the model effectively captures the multifaceted factors that influence car resale values, such as vehicle specifications, market trends, visual condition, and textual descriptions. This holistic approach provides a more comprehensive understanding of the complex dynamics of the used car market.

Through the use of multi-modal deep learning, various architectures, including Tab Net, LSTM, CNN, and BERT, collaborate to process different data types, ensuring a holistic approach to price prediction. The incorporation of explainability techniques, such as SHAP values [cite: a relevant SHAP paper], Grad-CAM [cite: a relevant Grad-CAM paper], and attention, enhances transparency, providing users with insights into the factors that drive price predictions. For instance, SHAP values can reveal how much each feature (e.g., mileage, condition, brand reputation) contributes to the final predicted price, allowing users to understand the model's reasoning. Grad-CAM can highlight the regions in an image that the CNN focuses on when assessing vehicle condition.

The project's structured workflow—encompassing data preprocessing, feature engineering, model ensembling, and deployment considerations—promotes scalability and real-world applicability. This solution has the potential to benefit a wide range of stakeholders, including individual car buyers, dealerships, online marketplaces, insurance companies, and financial institutions, by providing data-driven insights that support more informed decision-making in the used car market. Ultimately, this project demonstrates the transformative power of AI-driven pricing models in the automotive industry, with a focus on maintaining interpretability, building trust, and enhancing market efficiency.