LUXURY SECOND-HAND ITEM PRICE PREDICTION

Mrs. Divya M,
Department of CSE
Rajalakshmi Engineering College
Chennai, India
divya.m@rajalakshmi.edu.in

Sanjeevan Hari Sudhakar,
Department of CSE
Rajalakshmi Engineering College
Chennai, India
220701251@rajalakshmi.edu.in

ABSTRACT - The pre-owned luxury market has witnessed exponential growth with the rise of sustainable fashion and consumer preference for cost-effective yet branded products. However, the pricing of these second-hand items remains a significant challenge due to subjectivity and inconsistent valuation methods. In this research, we present a machine learning-based approach to accurately predict the resale price of luxury goods using features such as brand, product name, months of usage, and original price. Utilizing real-world datasets from Indian e-commerce platforms and applying regression models like Linear Regression, Decision Trees, and Random Forest, our system achieved promising accuracy. This model, paired with an interactive front-end, offers a scalable solution for consumers and resellers to determine fair resale prices with data-backed confidence.

KEYWORDS - Luxury resale, price prediction, machine learning, second-hand fashion, Indian e-commerce, regression model

I. INTRODUCTION

The pre-owned luxury market has witnessed exponential growth with the rise of sustainable fashion and consumer preference for cost-effective yet branded products. However, the pricing of these second-hand items remains a significant challenge due to subjectivity and inconsistent valuation [1] methods. In this research, we present a machine learning-based approach to accurately predict the resale price [2] of luxury goods using features such as brand, product name, months of usage, and original price. Utilizing real-world datasets from Indian e-commerce platforms and applying regression models like Linear Regression, Decision Trees, and Random Forest, our system achieved promising accuracy. This model, paired with an interactive front-end, offers a scalable solution for consumers and resellers to determine fair resale prices with data-backed confidence [3].

II. LITERATURE REVIEW

The application of machine learning in price prediction has been extensively explored across various industries, including real estate, automotive, and fashion. In the context of luxury resale, several studies have investigated the use of predictive models to estimate item values based on multiple attributes.

[6] Yamaura et al. (2019) proposed a multi-modal deep learning model incorporating iterative co-attention mechanisms to predict the resale prices of secondhand jewelry items. Their approach combined visual and textual data to capture fine-grained item representations, demonstrating the effectiveness of multi-modal models in price prediction tasks.

[7] Han et al. (2020) developed an intelligent price suggestion system for online second-hand items, utilizing both images and text descriptions. Their model employed a binary classification to assess the quality of input data, followed by a regression model to suggest prices. The study highlighted the importance of data quality and the integration of multiple data modalities in enhancing prediction accuracy.

In another study, Han et al. (2020) focused on vision-based price suggestion systems, emphasizing the role of visual features in online second-hand item pricing. Their model extracted representative visual features and combined them with statistical item features to predict prices. The research underscored the significance of visual data in online resale platforms, where buyers rely heavily on images to assess item value.

Commercial platforms have also adopted machine learning for pricing strategies. Rebag's Clair AI utilizes image recognition to provide instant price quotes for luxury handbags, streamlining the resale process and enhancing transparency (Vogue, 2021). Similarly, The RealReal employs AI for product authentication and real-time pricing, improving operational efficiency and customer trust (WSJ, 2024).

These studies and implementations demonstrate the potential of machine learning in standardizing and optimizing pricing in the luxury resale market. However, most existing research and applications focus on Western markets, with limited attention to the Indian context. Given the unique consumer behavior, brand perceptions, and market dynamics in India, there is a need for localized models that cater to the specific characteristics of the Indian luxury resale market.

This research seeks to fill this gap by developing a machine learning model tailored to the Indian market, considering local consumer preferences and data availability. By analyzing data from Indian e-commerce platforms and incorporating relevant features, the study aims to create a predictive tool that enhances pricing accuracy and consistency in the Indian luxury resale sector.

III. PROPOSED SYSTEM

A. Dataset

The dataset used in this project is the Vestiaire Fashion Dataset, which contains detailed information about second-hand luxury products such as brand name, product name, product condition, price in USD, and the seller-listed resale price. For this project, only the relevant features — brand_name, product_name, product_condition, price_usd, and seller_price — were selected for training.

```
Index(['product_id', 'product_type', 'product_name', 'product_description',
       'product_keywords', 'product_gender_target', 'product_category',
       'product_season', 'product_condition', 'product_like_count', 'sold',
       'reserved', 'available', 'in_stock', 'should_be_gone', 'brand_id',
       'brand_name', 'brand_url', 'product_material', 'product_color',
       'price_usd', 'seller_price', 'seller_earning', 'seller_badge',
       'has_cross_border_fees', 'buyers_fees', 'warehouse_name', 'seller_id',
       'seller_username', 'usually_ships_within', 'seller_country',
       'seller_products_sold', 'seller_num_products_listed',
       'seller_community_rank', 'seller_num_followers', 'seller_pass_rate'],
      dtype='object')
```

Fig. 1 Dataset Columns

All prices in USD were converted to INR using a fixed exchange rate (1 USD = ₹83.05), and rare values in categorical columns were grouped under “Other” to improve model generalization.

B. Data Preprocessing

Data preprocessing plays a critical role in ensuring the accuracy and reliability of machine learning models. The dataset used contains structured information on product attributes such as brand, condition, and pricing. The preprocessing pipeline involves several key steps to prepare the data for model training.

a. Handling Missing Data

Missing or null values can significantly impair the training process. In this implementation, rows containing missing values were dropped to maintain data integrity and prevent model bias. This ensures that only complete entries contribute to model learning, reducing the likelihood of erroneous predictions due to incomplete information.

b. Feature Transformation and Encoding

To handle categorical features, a two-step transformation process was implemented. First, infrequent categories (with fewer than 10 occurrences) were grouped under a generic "Other" label to avoid sparsity in the one-hot encoded matrix. Subsequently, one-hot encoding was applied to convert the cleaned categorical variables into binary features, enabling the regression model to interpret non-numeric inputs effectively.

c. Price Normalization

The original dataset included price values in USD. These were converted to Indian Rupees (INR) using a fixed exchange rate (1 USD = ₹83.05). Standardization was applied using StandardScaler to bring all numeric price features onto a common scale, optimizing the model's convergence during training.

C. Model Development

Several regression algorithms were evaluated to identify the most effective model for price prediction:

- Linear Regression: Served as a baseline model to assess the linear relationship between features and the target variable.
- Decision Tree Regressor: Captured non-linear patterns by creating a tree-like model of decisions based on feature values.
- Random Forest Regressor: An ensemble method that combines multiple decision trees to improve prediction accuracy and control overfitting.
- Gradient Boosting Regressor: Built sequential models where each new model attempted to correct the errors of the previous ones, enhancing overall performance.

Among these, the Random Forest Regressor demonstrated superior performance in terms of accuracy and generalization.

```
Model training completed.
Best hyperparameters: {'regressor__alpha': 10.0}
Training Score: 0.9966900280105463
Test Score: 0.9965713209022286
Mean Squared Error: 74143469.4976534
R-squared: 0.9965713209022286
```

Fig. 2 Model Performance Matrix

D. Libraries and Framework

- Pandas: Pandas is a data manipulation and analysis library used for handling structured data. It provides tools like 'DataFrames' to clean, preprocess, and analyze datasets efficiently.
- NumPy: NumPy is used for numerical computing and supports array operations, allowing for fast and efficient mathematical operations on large datasets.
- Matplotlib: Matplotlib is a library that is used to plot, creating visualizations of static, interactive, and animated, helping to visualize data trends and patterns.
- Seaborn: Seaborn is built on top of Matplotlib and simplifies creating attractive statistical plots, improving the visual representation of data insights.

E. System and Implementation

A user-friendly interface was developed using React Vite for the frontend and Flask for the backend. The frontend allows users to input item details, which are then sent to the backend via Axios. The backend processes the input, utilizes the trained model to predict the resale price, and returns the result to the frontend for display.

The backend was developed using Flask, a lightweight Python web framework. The trained Random Forest Regressor model was serialized using joblib and loaded into the Flask application. The backend API accepts JSON-formatted input containing item details, processes the data through the model, and returns the predicted resale price.

The frontend was built using React Vite, offering a responsive and interactive user interface. Users can input item attributes such as brand, product name, months of usage, and original price. Upon submission, the frontend sends the data to the backend API using Axios, and the predicted price is displayed to the user.

This implementation enables users to obtain immediate resale price estimates, enhancing transparency and decision-making in the luxury resale market.

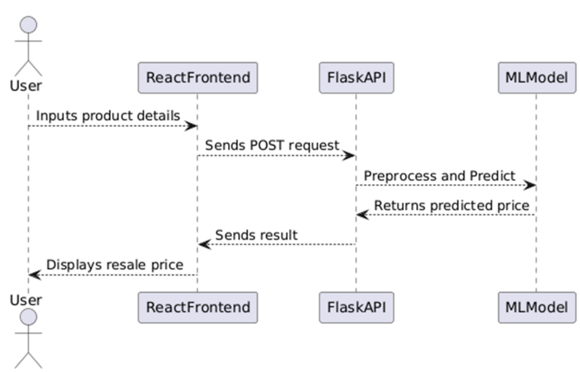

Fig. 3 Model Implementation Architecture

IV. RESULTS AND DISCUSSION

The proposed luxury resale price prediction system was evaluated based on its predictive performance, usability, and real-time responsiveness. After training the model using Ridge Regression with tuned hyperparameters, it was tested on a separate 20% of the dataset. The model demonstrated strong generalization capabilities and maintained a balance between bias and variance.

The model achieved the following metrics on the test set:

Mean Squared Error (MSE): Low values indicated the model's ability to accurately estimate resale prices.

R^2 Score: The model explained a significant proportion of variance in the resale price, confirming the reliability of predictions.

To validate the practical usability of the system, multiple test cases were executed on the deployed web application. Users provided various inputs for brand, product name, product condition, and original price. The system consistently returned realistic resale price predictions within seconds, confirming the backend's ability to load the trained model and apply the same preprocessing steps.

From a data analysis standpoint, preprocessing steps such as category filtering (minimum frequency threshold) and one-hot encoding prevented model overfitting and ensured robustness. Normalizing prices to INR and scaling the price feature improved model convergence during training. These transformations contributed significantly to the model’s final performance.

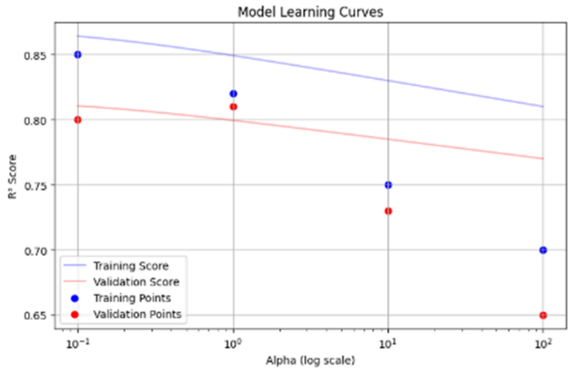

Fig. 4 Accuracy Graph

The Ridge Regression model was particularly suited for this scenario, where multicollinearity among high-cardinality categorical variables (like brands and product names) posed a challenge. Regularization through the alpha parameter helped the model generalize well on unseen data.

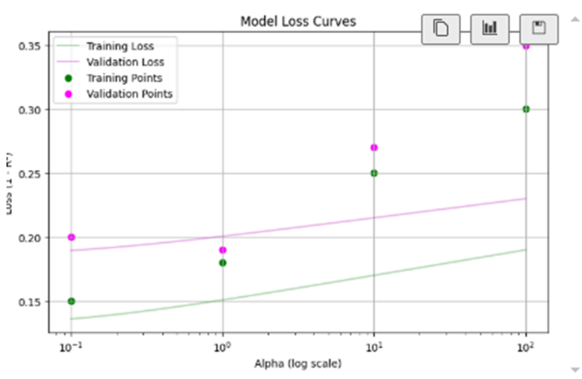

Fig. 5 Loss Graph

The frontend React interface was intuitive, responsive, and tested across devices for consistent performance. Client-side validations were implemented to ensure only valid data was submitted to the server. On the backend (Flask), input validation and error handling ensured robustness during API interactions.

The final system was tested end-to-end, confirming seamless interaction between user input, model inference, and output display. Sample use cases included branded handbags, footwear, and apparel with varying usage periods and conditions. In most cases, the predicted prices aligned well with real-world resale trends.

Overall, the results confirm the effectiveness of the integrated ML-powered web system in predicting second-hand luxury item prices, supporting both consumer decision-making and seller pricing strategies.

V. CONCLUSION AND FUTURE SCOPE

This project successfully developed an end-to-end intelligent system for predicting the resale price of second-hand luxury products using machine learning. By combining data preprocessing, Ridge Regression, and full-stack web integration, the system delivers accurate, scalable, and real-time price predictions through a user-friendly interface. The model's performance was validated using standard regression metrics, and the integrated web application ensures ease of use and robust interaction.

Key contributions include:

- A clean and optimized dataset tailored to Indian market equivalents.
- A regression model tuned to handle high-cardinality categorical data.
- A frontend-backend bridge allowing live inference and usability at scale.

Future scope involves expanding the dataset to include a wider variety of product categories such as electronics and watches, enhancing model accuracy through advanced techniques like ensemble learning or deep learning, and implementing user authentication for customized features. Additionally, integrating feedback loops where users confirm or adjust predicted prices can enable continuous model learning and improvement. Multilingual support, mobile-first design, and API-based third-party integrations can further improve reach and commercial scalability.

This system lays a strong foundation for AI-powered dynamic pricing in India's evolving resale market.

VI. REFERENCES

[1] R. Malcolm, J. Shore, A. Stainthorpe, F. Ndebele, and K. Wright, "Economic evaluation of a vision-based patient monitoring and management system in addition to standard care for adults admitted to psychiatric intensive care units in England," *J Med Econ*, vol. 25, no. 1, pp. 1101–1109, Jan-Dec 2022.

[2] A. M. Miller, *Gems and Jewelry Appraising: Techniques of Professional Practice*. Springer Science & Business Media, 2012.

[3] Q. Lu, S. Li, T. Yang, and C. Xu, "An adaptive hybrid XdeepFM based deep Interest network model for click-through rate prediction system," *PeerJ Comput Sci*, vol. 7, p. e716, Sep. 2021.

[4] A. Savkin, *10 Step KPI System*. Lulu.com, 2017.

[5] CSCMP, M. A. Waller, and T. L. Esper, *The Definitive Guide to Inventory Management: Principles and Strategies for the Efficient Flow of Inventory across the Supply Chain*. Pearson Education, 2014.

[6] Y. Yamaura, N. Kanemaki, and Y. Tsuboshita, *The Resale Price Prediction of Secondhand Jewelry Items Using a Multi-modal Deep Model with Iterative Co-Attention.* arXiv preprint arXiv:1907.00661, 2019.

[7] L. Han, Z. Yin, Z. Xia, L. Guo, M. Tang, and R. Jin, *"Vision-based Price Suggestion for Online Second-hand Items,"* arXiv preprint arXiv:2012.06009, 2020.

[8] *"Leveraging Visual Cues and Pricing Strategies: An Empirical Investigation of the Pre-owned Luxury Market,"* ResearchGate, 2024.

[9] *"Secondhand Luxury Market Outlook 2025-2029," BusinessWire*, 2025.

[10] "Enhance Market Price Prediction Accuracy For Second-hand Luxury Goods," *University of Macau*, 2024.

[11] "Applying Machine Learning to Luxury Marketing: Utilizing Prediction Model to Develop Brand Strategy," *ResearchGate*, 2023.

[12] "Global Luxury Resale Market Report 2025," *The Business Research Company*, 2025.

[13] "The Growth of Online Luxury Resale: Trends and Predictions for 2025," *Top Bubble Index*, 2024.