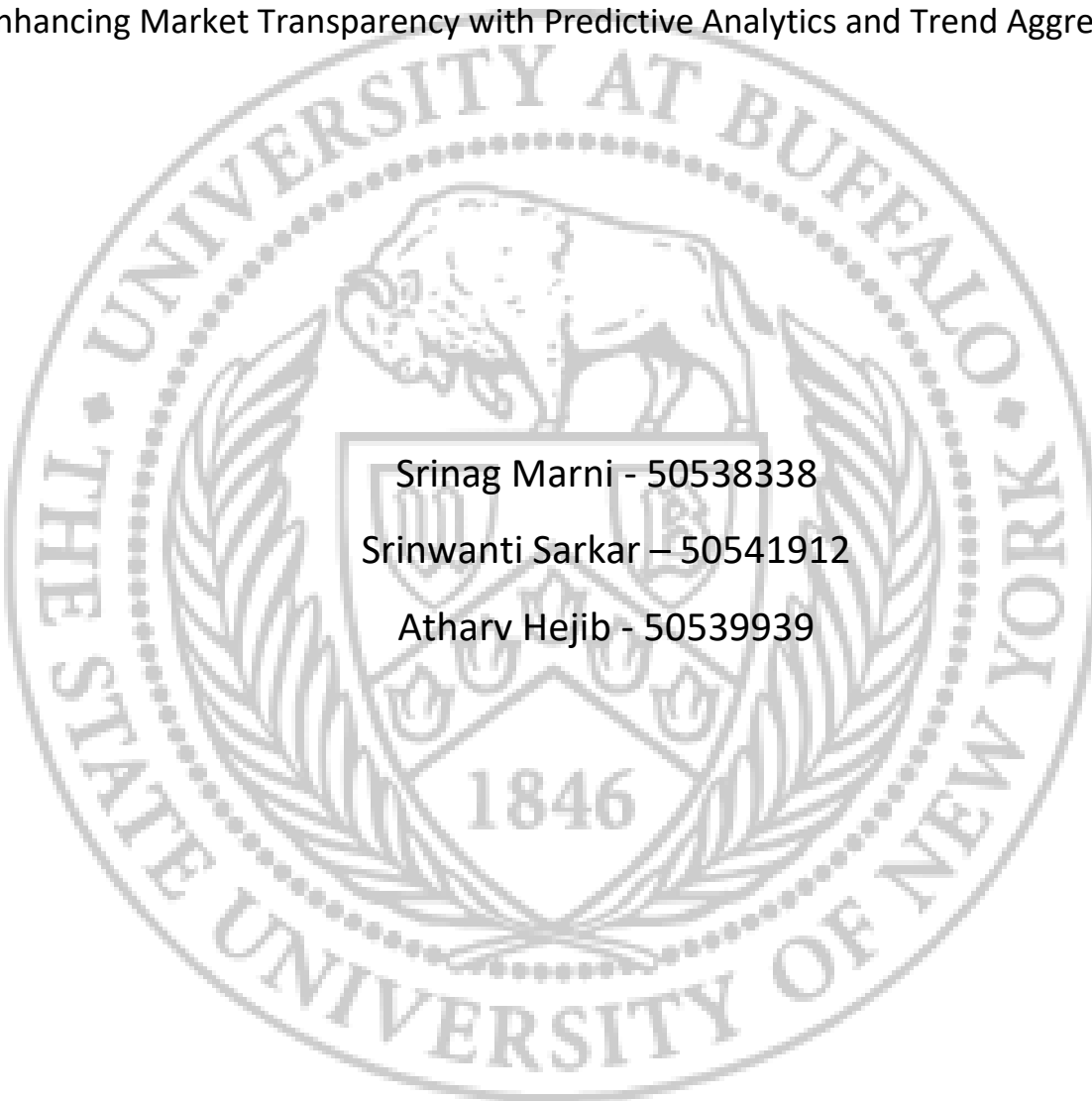


Price Master: Unveiling Car Price Trends with Data Analysis

Enhancing Market Transparency with Predictive Analytics and Trend Aggregation



Srinag Marni - 50538338

Srinwanti Sarkar – 50541912

Atharv Hejib - 50539939

A Report Submitted in partial fulfillment of the requirements for the
Engineering Science (Data Science) MS

DEPARTMENT OF
Institute for Artificial Intelligence and Data Science
School of Engineering and Applied Sciences
University at Buffalo

INTRODUCTION

The primary objective of our project is to develop a user interface that allows users to input their own dataset for automatic analysis. This interface will leverage the machine learning and statistical models developed in Phase 2 to provide users with insights into the factors influencing car resale prices.

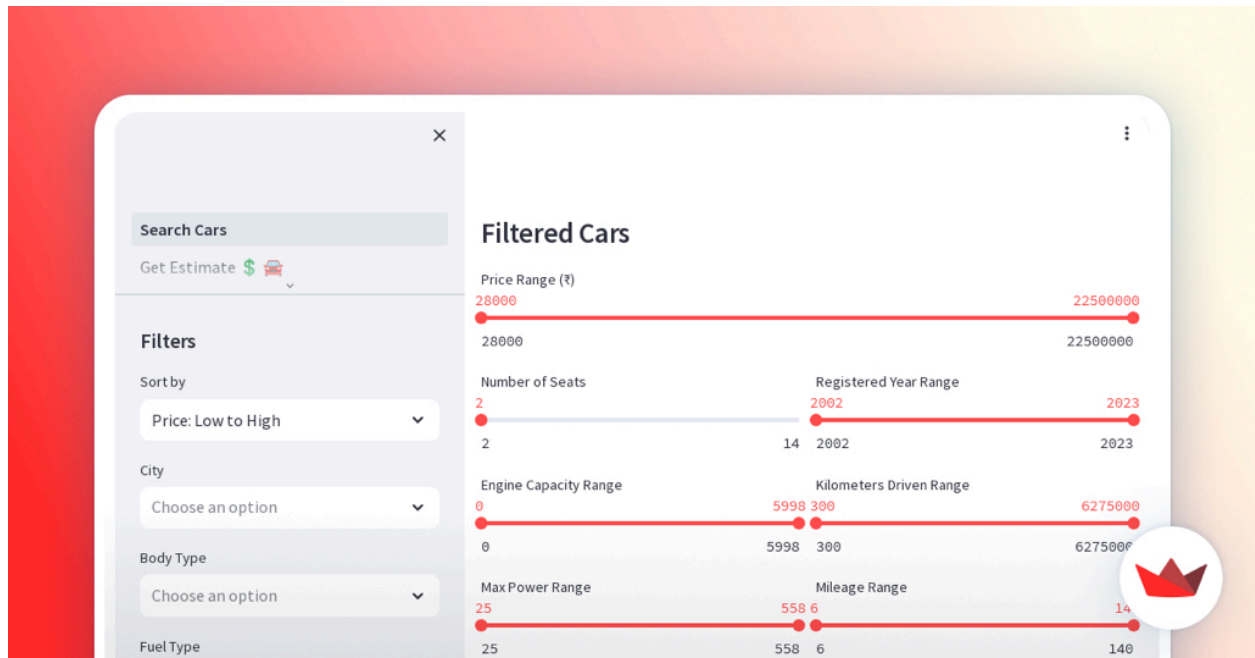
By providing users with access to our advanced modeling techniques through an intuitive interface, our data product aims to simplify the process of gaining insights into car resale prices. Whether they are car dealers, analysts, or enthusiasts, our product will enable users to make informed decisions based on data-driven insights.

RECAP

Phase 1 of our project marked the start of our journey towards understanding the dynamics of car resale prices. During this phase, we undertook comprehensive data collection and preprocessing activities. Our primary objectives in Phase 1 included sourcing relevant datasets encompassing a wide array of car attributes, such as mileage, age, brand, model, and condition. Additionally, we conducted rigorous data cleaning and preprocessing to ensure the integrity and quality of the dataset. We standardized variable formats, handled missing values, and resolved inconsistencies, thereby preparing a pristine dataset for further analysis. Furthermore, we conducted exploratory data analysis (EDA) to gain preliminary insights into the distribution, correlation, and characteristics of the variables.

Phase 2 involved applying machine learning and statistical modeling algorithms to analyze a preprocessed dataset on car attributes and resale prices. The goal was to identify factors influencing resale prices and answer key questions related to car pricing dynamics. Various algorithms, including Linear Regression, Ridge Regression, Decision Tree Regression, Gradient Boosting, Random Forest Regression, and XGBoost Regressor, were used to learn insights about resale price trends. Overall, Phase 2 highlighted the importance of algorithm selection and evaluation in gaining knowledge for pricing strategies in the automotive industry.

DATA PRODUCT DESCRIPTION



The "Price Master" application, designed with Streamlit, serves as a comprehensive platform for estimating and analyzing the resale prices of cars. This tool is tailored for individual car owners looking to understand the potential market value of their vehicles, businesses involved in the used car market, and consumers actively searching for cars to purchase. The application offers a dual functionality: direct price estimation for individual cars and batch processing for multiple entries via CSV upload.

At the core of "Price Master" is a machine learning model, specifically an XGBoost regressor, which has been trained on a vast dataset of car resale values. This model leverages various car attributes such as the make and model, age, mileage, and other specifications to predict resale prices with precision. Users can input the details of a car through a user-friendly form, which includes fields for the car's make, model year, mileage, and other relevant attributes. Upon submission, the model processes this data to provide an immediate resale price estimate.

For those searching for cars, the application includes a variety of filters and sorting options, allowing users to sift through a large database of cars based on criteria such as price, mileage, and year of registration. This feature is particularly valuable for potential buyers, as it helps them quickly find cars that match their specific preferences and budget, facilitating a more targeted and efficient shopping experience.

Additionally, the batch estimation feature is a significant advantage for businesses. It allows the uploading of a CSV file containing multiple car details, which the application processes to append estimated prices. This can greatly aid businesses in evaluating their inventory or in making bulk purchasing decisions.

Overall, the design and functionality of "Price Master" reflect a deep understanding of its diverse user base, offering both simplicity in individual estimations and efficiency in handling bulk data. The platform is a versatile tool in the car resale market, catering to a wide range of needs from price evaluation to car shopping.

METHODOLOGY

Development Approach

Integration of Models:

Incorporated the machine learning models developed in Phase 2, specifically Linear Regression, Ridge Regression, Decision Tree, Random Forest, Gradient Boosting, and XGBoost, to predict car resale prices based on user inputs.

Model Utilization:

The application dynamically selects the most suitable model based on the complexity of the input data. For simpler queries, simpler models like Linear Regression might be used, whereas for more complex queries, ensemble methods like Random Forest or XGBoost are employed.

User Interaction Design:

Focused on creating a straightforward workflow where users can easily input data and receive outputs without needing to understand the underlying complexity.

Model Implementation Details

Parameter Tuning: Utilized techniques like grid search and cross-validation primarily with complex models like XGBoost and Random Forest to fine-tune hyperparameters for optimal performance.

Feature Importance: Integrated insights from models to highlight which features most significantly impact the predicted resale price, enhancing user understanding and trust in the predictions.

CODE OVERVIEW

Usage of XGBoost in "Price Master"

XGBoost (eXtreme Gradient Boosting) was chosen for its efficiency and superior performance, particularly in scenarios involving datasets with intricate patterns and diverse feature types. Here's how XGBoost was implemented:

1. Model Training:

- XGBoost was trained on the preprocessed dataset where features like age, mileage, engine capacity, and other car attributes were used as predictors for the car resale prices.

2. Feature Importance:

- XGBoost provides an in-built method to evaluate feature importance, which was utilized to inform users about which factors are most influential in determining resale prices. This helps in transparency and educates the user on what affects their car's value.

3. Predictive Performance:

- The model's high R-squared value indicated that it could explain a significant proportion of variance in the resale prices, making it reliable for practical use in the data product.

4. Parameter Tuning of XGBoost

Tuning XGBoost involved adjusting several hyperparameters to optimize the model's performance:

- **n_estimators**: Specifies the number of gradient boosted trees. Too few might underfit, while too many can lead to overfitting.
- **learning_rate**: Controls the contribution of each tree. Lowering the learning rate requires more trees but can lead to better ensemble predictions.
- **max_depth**: Sets the maximum depth of each tree. Deeper trees can learn more detailed data specifics, enhancing performance but increasing the risk of overfitting.
- **subsample and colsample_bytree**: These parameters manage the fraction of samples and features used per tree, introducing randomness and reducing overfitting.
- **reg_alpha and reg_lambda**: These are regularization parameters that help reduce model complexity and enhance performance on unseen data.

Integration in Data Product

In the "Price Master" application, XGBoost is integrated to serve as the primary predictive model, especially for complex user queries where accuracy is critical:

Dynamic Model Selection: Depending on the complexity and nature of the user input, the application decides whether to employ XGBoost or other models. XGBoost is preferred for detailed and comprehensive inputs where the user expects highly accurate predictions.

User Interaction: The application allows users to see how different inputs would affect the predicted resale value, leveraging the interpretative benefits of XGBoost's feature importance.

The choice of XGBoost for the "Price Master" application is justified by its robust performance across various metrics and its suitability for the dataset used. The careful tuning of its parameters ensures that the model is both accurate and generalizable, making it an excellent tool for predicting car resale prices in a real-world application. This integration helps make "Price Master" a valuable tool for users looking to get precise resale value estimations based on a wide array of vehicle features.

MODEL TUNING AND EVALUATION

XGBoost was employed as the main algorithm for regression tasks, given its efficiency and ability to handle complex relationships within the data. The default hyperparameters of the XGBRegressor class were initially used, but in a real-world scenario, further optimization could be achieved through hyperparameter tuning techniques like grid search or randomized search.

Hyperparameters

- **Learning Rate:** The default learning rate of 0.1 was used, controlling the contribution of each tree to the ensemble.
- **Max Depth:** The maximum depth of each tree was set to 3 by default, limiting the complexity of the individual trees.
- **Number of Estimators:** The number of boosting rounds was set to 100 by default, specifying the number of trees to build.

Model Performance

- **Mean Absolute Error (MAE):** The XGBoost regression model achieved an MAE of 107590.57 on the test data, indicating the average magnitude of the errors in predicting resale prices.
- **Root Mean Squared Error (RMSE):** The RMSE of 238746.93 suggests the square root of the average squared differences between predicted and actual resale prices.

- **R-squared Value:** The high R-squared value of 0.9296 indicates that the model explains a significant proportion of the variance in the target variable, implying strong predictive performance.

Statistical Significance

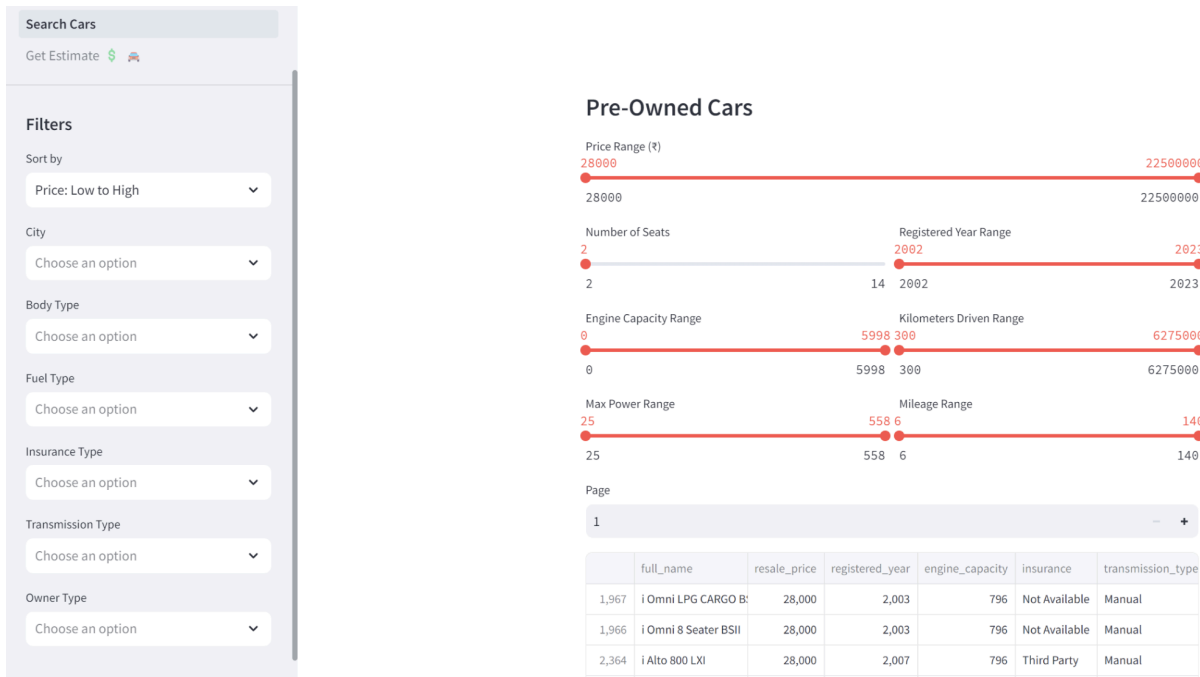
- **F-Statistic:** The high F-Statistic value of 3307.99 suggests that the overall model is statistically significant, indicating that the features collectively have a strong relationship with the target variable.

Insights

- **Feature Importance:** XGBoost provides feature importance scores, aiding in feature selection and interpretation. This highlights the most influential features in predicting resale prices, such as car age, mileage, and engine capacity.
- **Complex Data Patterns:** XGBoost excels in handling complex data patterns inherent in the car resale dataset, effectively learning and exploiting intricate relationships for improved predictive performance.

In conclusion, XGBoost regression proves to be effective for predicting resale prices in this context, achieving high accuracy and performance. Further optimization through hyperparameter tuning and feature engineering could potentially enhance the model's accuracy and generalizability.

APPLICATION OVERVIEW





The UI allows users to interact with the data product in a user-friendly manner. Users can filter cars based on various criteria such as city, body type, fuel type, insurance type, transmission type, and owner type. The filtered data is displayed in a tabular format, allowing users to easily compare different car options.

Additionally, users can input specific car details such as car name, owner type, registered year, engine capacity, kilometers driven, maximum power, seats, mileage, insurance type, transmission type, fuel type, and body type to get an estimated price for the car.

At the heart of our project lies a robust XGBoost model, meticulously trained to analyze a preprocessed dataset brimming with car attributes and resale prices. We have meticulously approached this problem from a user perspective, aiming to alleviate the pain points faced by both car buyers and sellers in navigating this complex market landscape.

×

Search Cars


Get Estimate  

Try Our Batch Estimator

Upload CSV File

Drag and drop file here
Limit 200MB per file • CSV

Browse files



Download Sample Template

Car Price Estimator

Enter Car Details to get Estimated price

Car Name	Owner Type
<input type="text"/>	First Owner
Registered Year	Engine Capacity (cc)
2016	814.00
Kilometers Driven	Maximum Power (bhp)
68336.40	55.00
Number of Seats	Mileage
5	21.00
Insurance	Transmission Type
Zero Dep	Manual
Fuel Type	Body Type
Petrol	Hatchback
Get Price	

For individuals seeking their dream car, our search cards feature provides a seamless experience, allowing users to filter through a myriad of options based on their specific preferences and requirements. Whether it's meeting a certain price point or finding a vehicle with particular specifications, we have them covered.

But our platform doesn't stop there. For business owners, we offer invaluable insights into market trends and demand fluctuations. By analyzing which segments are underserved or where there's a scarcity of certain car models, our platform empowers businesses to make informed decisions about their inventory, ensuring they stock up on the right cars at the right time.

One of the standout features of our website is the 'Estimator' tab, where users looking to sell their cars can obtain an accurate prediction of their vehicle's resale value. This not only provides sellers with a clearer understanding of their car's worth but also enables businesses to evaluate their inventory's value and make strategic decisions accordingly.

Furthermore, the data product includes a batch estimator that allows users to upload a CSV file containing multiple car records. The batch estimator then adds a predicted price column to the CSV file, which users can download for further analysis.

Overall, the data product provides a comprehensive solution for estimating car prices and gaining insights into the automotive market, offering valuable tools for buyers, sellers, and analysts in the industry.

RECOMMENDATIONS

Suggestions for Users

- Price Comparison/Decision Making: Users can benefit from the data product by comparing prices of different cars based on their specifications and market trends. It can help users make informed decisions when buying or selling a car by providing estimated prices based on various factors.
- Customized Search: Users can filter cars based on their preferences such as price range, mileage, and other specifications to find the most suitable options.
- Market Understanding: Analyzing data reveals trends and customer preferences, aiding market adaptation.
- Competitive Edge: ML models optimize pricing, providing a competitive advantage in the market.
- Customer Segmentation: Identify customer segments to tailor marketing and product strategies.

Problem Solving

- Price Estimation: The data product can help solve the problem of estimating resale prices of cars, providing a valuable tool for both buyers and sellers in the automotive market.
- Market Insights: Users can gain insights into market trends and pricing dynamics, helping them understand the value of their cars in the current market.
- Risk Management: Predictive models help manage market risks and uncertainties effectively.
- Innovation Opportunities: Uncover market gaps, inspiring innovation in car design and features.
- Investor Confidence: Accurate predictions instill investor confidence, guiding strategic decisions

Ideas for Extension

- Enhanced Features: Integrate additional features such as predictive maintenance suggestions, insurance cost estimations, or personalized car recommendations.
- User Feedback: Incorporate a feedback mechanism to improve the accuracy and usability of the data product based on user inputs.
- Data Expansion: Extend the dataset to include more car models, years, and regions for a more comprehensive analysis.
- Collaborate with existing tools like Facebook Marketplace , Carmax , Quiker , Cardeko to add more data and expand our understanding
- Additionally, we can later offer a single platform that includes an extension to post listings to all marketplaces for used cars.

By implementing these suggestions, the data product can become an indispensable tool for anyone involved in the automotive market, providing valuable insights and facilitating informed decision-making.

CONCLUSION

Throughout the development process, we've encountered various challenges and learning opportunities, from data cleaning to model integration and interface design. Each phase has contributed to our project's success, allowing us to refine our approach and create a robust and functional data product.

Our project is all about simplifying the car resale experience. We've combined smart technology with user-friendly design to create a platform that's easy to use and packed with insights. Whether you're buying or selling, our platform has you covered with accurate price estimates and market trends.

With features like the 'Estimator' tab and batch processing, we've made it simple for users to get the information they need, fast. And we're not stopping there – we're always looking for ways to improve and add new features to make your experience even better.

Our platform not only simplifies the process of estimating resale prices but also empowers users with actionable market intelligence, fostering informed decision-making. As we navigate the future, we remain committed to innovation, continually enhancing our platform to meet the evolving needs of our users and driving positive change in the automotive industry.

Looking ahead, there are several areas for future improvement and expansion. We aim to enhance the user interface with additional features and further improve the accuracy of our machine-learning models. Deploying the data product commercially and expanding it into a comprehensive car marketplace platform are also key considerations for the future.

REFERENCES

- <https://bootcamp.uxdesign.cc/ux-case-study-website-to-buy-sell-used-cars-fe4fa2ba31dd>
- <https://blog.streamlit.io/crafting-a-dashboard-app-in-python-using-streamlit/>
- https://xgboost.readthedocs.io/en/stable/python/python_api.html
- https://seaborn.pydata.org/tutorial/color_palettes.html
- **Dataset:** <https://www.kaggle.com/datasets/rahulmenon1758/car-resale-prices>

Link to Code:

🔗 Price_Master_pahse_3.ipynb

🔗 Price_Master_Phase_2.ipynb

🔗 Price_Master_Phase_1.ipynb

Website Link: [Price Master](#)

Repo: [Price Master Repo](#)

PROJECT DETAILS

Group Members:

- Srinag Marni
- Atharv Hejib
- Srinwanti Sarkar

Our deepest gratitude to **Prof. Shamsad Parvin** for invaluable guidance.

Special acknowledgement to **Jaskirat Singh** for pivotal support in data processing/model Building, benefiting all.