# Used Car Price Prediction

## Overview

This project aims to predict the prices of used cars using various machine learning models. The dataset includes several features related to the car's specifications, such as year, manufacturer, odometer, fuel type, and transmission. The project focuses on building a robust predictive model to estimate car prices accurately based on these features.

## Purpose

The used car market has seen substantial growth in recent years due to various economic factors, including increased demand for affordable transportation and supply chain disruptions affecting new car production. Accurate price prediction can help sellers price their vehicles competitively and assist buyers in making informed purchasing decisions. The goal of this project is to explore different regression models and select the most effective one for predicting used car prices.

## Methodology

The project follows a systematic approach to data analysis and modeling:

1. \*\*Data Preparation\*\*: The dataset was cleaned to handle missing values, remove outliers, and encode categorical variables.

2. \*\*Feature Engineering\*\*: New features were created, such as age categories and log-transformed variables, to improve model performance.

3. \*\*Model Selection\*\*: Several regression models were tested, including:

- Linear Regression

- Ridge Regression

- Lasso Regression (with cross-validation for optimal alpha)

- Decision Tree Regression

- Random Forest Regression

4. \*\*Evaluation and Analysis\*\*: Each model was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R²) metrics to determine its accuracy and effectiveness.

## Key Findings

After evaluating multiple models, the \*\*Random Forest Regression\*\* model was chosen as the best performer due to its superior accuracy and ability to handle complex, non-linear relationships within the data. The Random Forest model provided the lowest MAE and MSE and a significantly higher R-squared value compared to other models, indicating a better fit for the data.

## Files <https://drive.google.com/file/d/1B-TkRd24702dgLkJzShTZ5-6UvN1tpDo/view?usp=drive_link>

- `FINAL\_Used\_Car\_Price\_Prediction\_Practical\_Application\_II.ipynb`: The Jupyter Notebook containing the entire workflow, including data preparation, feature engineering, model training, and evaluation.

- `FINAL\_Used\_Car\_Price\_Prediction\_Practical\_Application\_II.html`: The HTML version of the notebook for easy viewing.

## How to View the Project

To view the project in its entirety:

1. \*\*HTML Version\*\*: Open the `FINAL\_Used\_Car\_Price\_Prediction\_Practical\_Application\_II.html` file in any web browser to explore the analysis and results.

2. \*\*Jupyter Notebook\*\*: Download the `FINAL\_Used\_Car\_Price\_Prediction\_Practical\_Application\_II.ipynb` file and open it using Jupyter Notebook to run the code cells interactively.

## Conclusion

The project successfully demonstrated the application of machine learning techniques to predict used car prices accurately. The chosen model, Random Forest Regression, outperformed other models and is recommended for further analysis and deployment. Future work could involve incorporating more features, such as geographical data or market trends, to enhance model accuracy.

## How to Contribute

Contributions are welcome! If you have suggestions for improvements, new features, or additional analysis, please feel free to fork the repository and create a pull request.

## License

This project is licensed under the UCB License - see the [LICENSE](LICENSE) file for details.

## Acknowledgments

Special thanks to the contributors and data providers who made this project possible.