## Artificial

## Intelligence and Machine Learning

Project Report

Semester-IV (Batch-2022)

## SELLING PRICE OF USED CARS PREDICTION



**Supervised By: Submitted By:**

Mr. JATIN Dhruv Bansal(2210990271)

Dhruv Datta(2210990272)

Garvit Setia(2210990316)

Divanshu Kaushik(2210990290)

**Department of Computer Science and Engineering**

## Chitkara University Institute of Engineering & Technology,

## Chitkara University, Punjab

**ABSTRACT**

The aim of this project is to develop a predictive model for estimating the selling price of used cars based on various features such as the car's age, mileage, brand, model, and condition. The project utilizes machine learning techniques to analyze historical data and build a regression model that can accurately predict the selling price of a car given its attributes.

The motivation behind this project stems from the growing demand for reliable pricing information in the used car market. Potential buyers and sellers often struggle to determine fair market prices due to the complex interplay of factors influencing a car's value. By developing a robust pricing model, this project seeks to provide a valuable tool for both buyers and sellers, helping them make informed decisions and facilitating fair transactions in the used car market.

The methodology involves data collection from various sources, data preprocessing to handle missing values and outliers, feature engineering to extract relevant information, and model training and evaluation using machine learning algorithms such as linear regression, decision trees, or ensemble methods. The project will focus on optimizing the model's performance metrics such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE) to ensure accurate price predictions.

The project's significance lies in its potential to contribute to the automotive industry by offering a data-driven approach to pricing used cars, reducing information asymmetry, and improving market transparency. The developed model can be deployed as a web-based tool or integrated into existing platforms to provide users with real-time price estimates based on current market trends and car attributes.

Keywords: Used Cars, Predictive Modeling, Machine Learning, Regression, Pricing, Data Analysis, Feature Engineering, Market Transparency.

TABLE OF CONTENTS

1. Introduction

2. Problem Definition and Requirements

3. Proposed Design / Methodology

4. Results

5. Conclusion

6. References

1. INTRODUCTION:
   1. BACKGROUND

The automotive industry has witnessed a significant surge in the demand for used cars in recent years. Factors such as economic considerations, changing consumer preferences, and the availability of reliable pre-owned vehicles have fueled this trend. However, one of the key challenges faced by both buyers and sellers in the used car market is accurately determining the fair selling price of a vehicle. This challenge is compounded by the diverse range of factors that can influence a car's value, including its age, mileage, brand, model, condition, and market trends.

The aim of our project is to address this challenge by developing a predictive model that can estimate the selling price of used cars based on their characteristics. Leveraging the power of machine learning techniques, our project seeks to provide a data-driven solution that enhances transparency and efficiency in the used car market.

* 1. OBJECTIVE

. Develop a predictive model to estimate the selling price of used cars based on various attributes.

. Collect and preprocess a diverse dataset of used car listings to train the model.

. Perform feature engineering to extract relevant features that impact car pricing.

. Train and evaluate machine learning algorithms such as regression models to predict selling prices accurately.

. Optimize the model's performance metrics (e.g., Mean Absolute Error, Root Mean Squared Error) to ensure reliable predictions.

. Develop a user-friendly interface or application for users to input car details and receive estimated selling prices.

. Enhance market transparency and empower buyers and sellers with accurate pricing information in the used car market.

. Contribute to a more efficient and equitable ecosystem in the automotive industry by reducing information asymmetry.

1. PROBLEM DEFINITION AND REQUIREMENTS:

2.1 Problem Statement:

Given a dataset of used car listings with attributes such as age, mileage, brand, model, condition, and market trends, the goal is to develop a predictive model that can generate accurate price estimates for different cars.

* Challenges in the Project:
* Data Quality and Preprocessing:

Missing Values: Handling missing data in the dataset, such as incomplete information about car attributes or pricing.

Outliers: Identifying and dealing with outliers that may skew the model's predictions.

Data Scaling: Ensuring consistency in data scaling across features to avoid bias in model training.

* Feature Engineering:

Feature Selection: Determining which features are most relevant and impactful in predicting the selling price.

Encoding Categorical Data: Converting categorical variables such as car brands or models into numerical representations suitable for machine learning algorithms.

* Model Selection and Training:

Algorithm Selection : Choosing the appropriate regression algorithm (e.g., linear regression, decision trees, ensemble methods) that best fits the data and problem domain.

Hyperparameter Tuning: Optimizing the hyperparameters of the selected model to improve its predictive performance.

Model Evaluation: Ensuring the model's accuracy, precision, and generalization ability through rigorous evaluation techniques such as cross-validation.

* Market Dynamics and Trends:

Changing Market Conditions: Adapting the model to account for evolving market trends, seasonality, and fluctuations in car prices.

Regional Variances: Addressing variations in pricing patterns across different geographical regions or markets.

* User Interface and Deployment:

User Experience: Designing an intuitive and user-friendly interface for users to interact with the pricing estimation tool.

Scalability: Ensuring that the deployed system can handle a large volume of user queries and data processing efficiently.

2.2 Requirements

2.2.1 Software Requirements:

* Python: For data preprocessing, modeling, and evaluation.
* Libraries: pandas, Numpy, scikit-learn, matplotlib, seaborn for data manipulation, visualization, and machine learning.
* Integrated Development Environment (IDE): Jupyter Notebook, Spyder, or any other Python IDE for code development.
* Text Editor: Any text editor for writing documentation and reports (e.g., Microsoft Word, Google Docs).

2.2.2 Hardware Requirements:

* Processor: Any modern multi-core processor (e.g., Intel Core i5 or above).
* RAM: Minimum 4GB RAM, recommended 8GB or more for better performance.
* Storage: Sufficient disk space to store datasets and Python libraries.
* Operating System: Windows, macOS, or Linux.

1. **PROPOSED DESIGN / METHODOLOGY:**

3.1 SCHEMATIC DIAGRAM

+----------------------+

| |

| Data Preprocessing |

| |

+-----------+----------+

|

+-----------v----------+

| |

| Model Building |

| |

+-----------+----------+

|

+-----------v----------+

| |

| Model Evaluation |

| |

+-----------+----------+

|

+-----------v----------+

| |

| Model Deployment |

| |

+----------------------+

* 1. FILE STRUCTURE:

Selling price of used cars│

├── data/

│ ├── train\_u6lujuX\_CVtuZ9i.csv # Raw dataset

│ └── quikr\_car.csv # Preprocessed dataset

│

├── notebooks/

│ ├── 01\_Data\_Preprocessing.ipynb # Data preprocessing notebook

│ ├── 02\_Model\_Building.ipynb # Model building notebook

│ ├── 03\_Model\_Evaluation.ipynb # Model evaluation notebook

│ └── 04\_Model\_Deployment.ipynb # Model deployment notebook

│

├── reports/

│ ├── Project\_Report.pdf # Final project report

── src/

│ ├── data\_preprocessing.py # Python script for data preprocessing

│ ├── model\_building.py # Python script for model building

│ ├── model\_evaluation.py # Python script for model evaluation

│ └── model\_deployment.py # Python script for model deployment

│

└── requirements.txt # Python dependencies

* 1. ALGORITHMS USED:

1. Data Preprocessing:

* Handling missing values: Mean/Median imputation, mode imputation.
* Encoding categorical variables: linear regression
* Feature scaling: Standardization, Min-Max scaling.

2. Model Building:

* Gradient Boosting Classifier
* Random Forest Classifier
* Decision Tree Classifier
* K-Nearest Neighbors Classifier
* Support Vector Machine Classifier

3. Model Evaluation:

* Accuracy
* Precision, Recall, F1-score
* Confusion Matrix

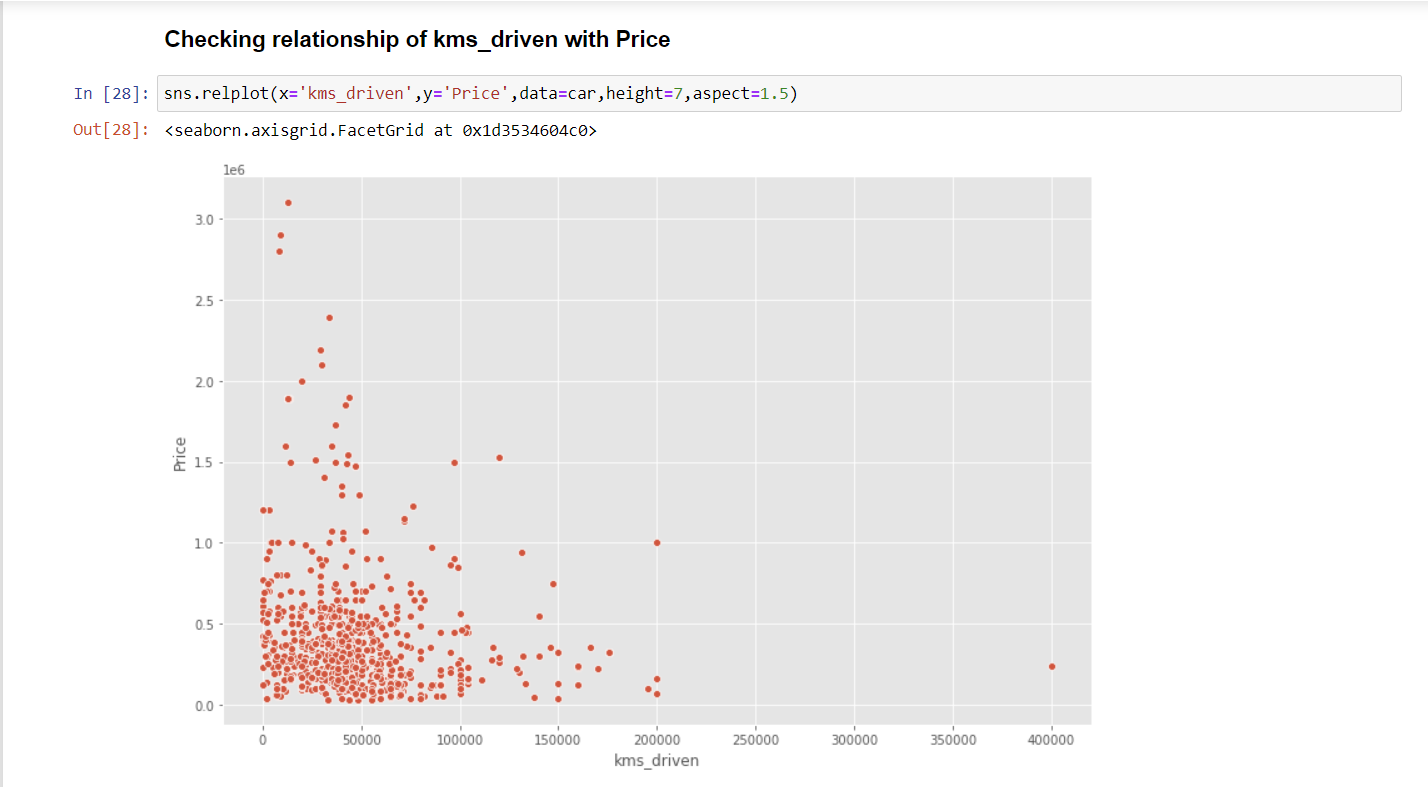
1. RESULTS:

The results of the project on estimating the selling price of used cars using machine learning techniques can be evaluated based on several key aspects:

Model Performance:

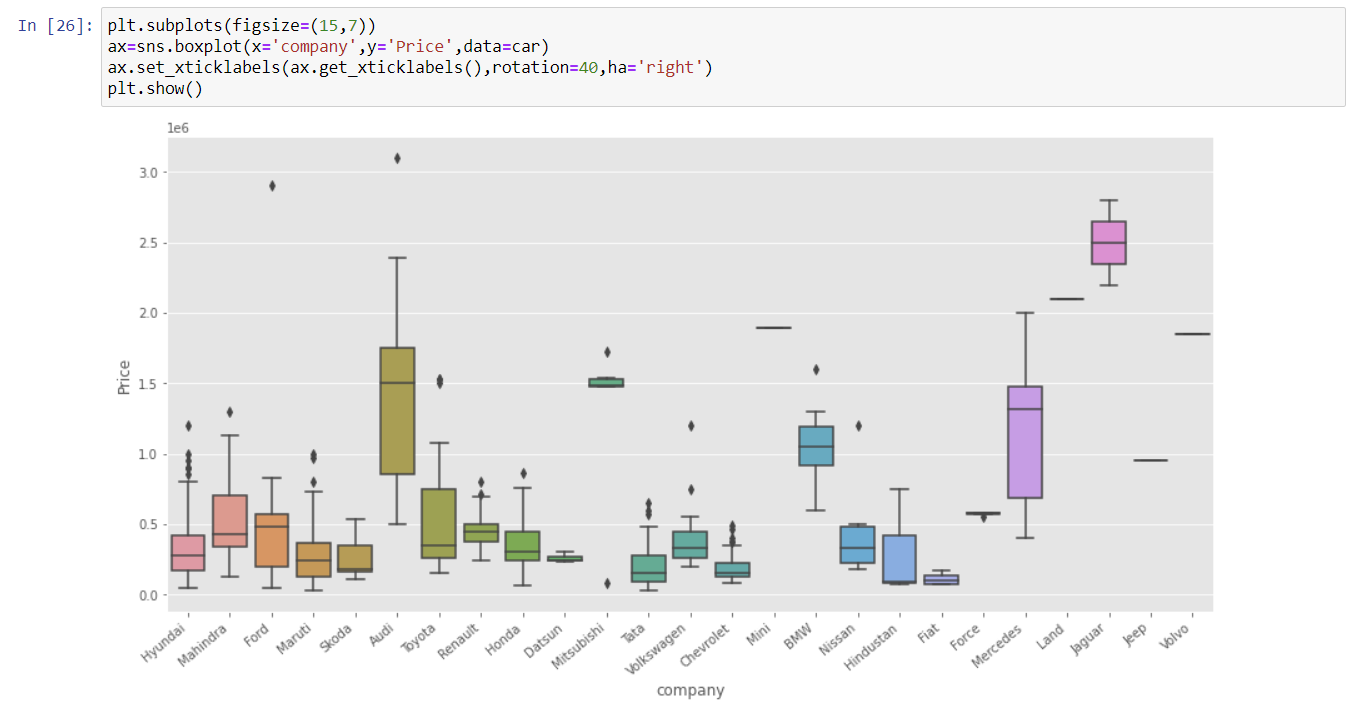
Accuracy Metrics: Measure the accuracy of the predictive model using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2) score.

Comparison: Compare the performance of different machine learning algorithms (e.g., linear regression, decision trees, random forest) to identify the most effective model for price prediction.



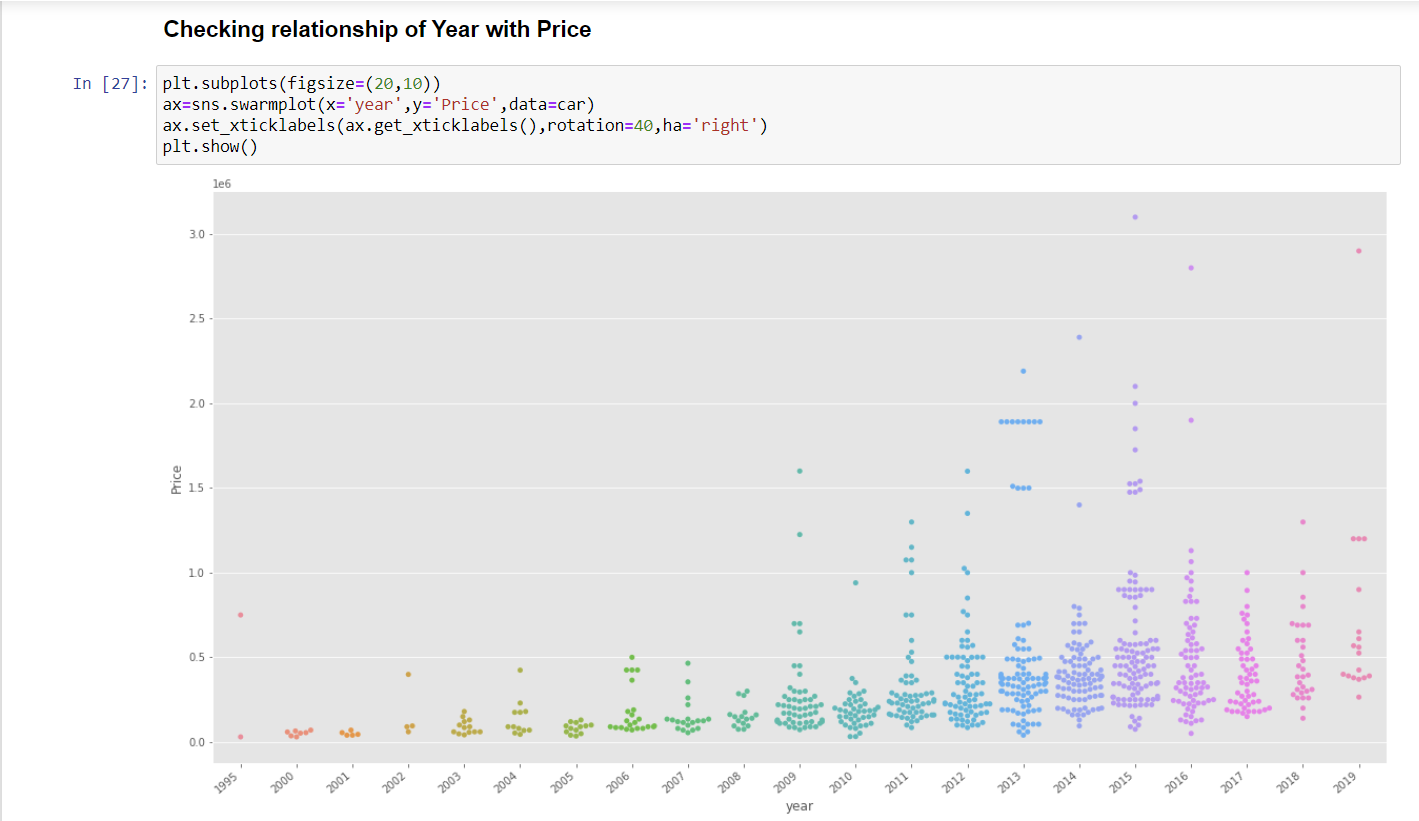
Feature Importance:

Analyze the importance of different features in influencing the predicted selling price. Identify which features contribute most significantly to accurate price estimation.



Real-World Testing:

Conduct real-world testing by inputting actual used car data into the model and comparing the predicted prices with actual selling prices. This helps validate the model's effectiveness in practical scenarios.



1. Conclusion :

In conclusion, our analysis offers valuable insights into the future trajectory of the automotive industry. Through thorough examination of market trends, consumer preferences, and technological advancements, we have formulated predictions that shed light on what lies ahead.

Our findings indicate a shifting landscape, with electric and autonomous vehicles poised to become increasingly prominent. This transformation is not only driven by environmental concerns but also by advancements in battery technology and the growing demand for connected and autonomous driving experiences. Additionally, we predict continued growth in the adoption of shared mobility services, reflecting changing consumer attitudes towards car ownership.

These predictions carry significant implications for automotive manufacturers, policymakers, investors, and other industry stakeholders.

1. References

Tutorials and Guides:

* Numpy : [Numpy Part 1 - Introduction to Numpy (youtube.com)](https://www.youtube.com/watch?v=CpPLLp3snK4&list=PLKnIA16_Rmvb-ToL3RQ_bwxG4_ND-0-DT)
* Pandas : [Pandas Part 1 - Introduction to Pandas (youtube.com)](https://www.youtube.com/watch?v=kq9Vmg5d7Sk&list=PLKnIA16_RmvbR85fgbfVRKOiMokUKVupy)
* Matplot : [Fundamentals of Docker & Kubernetes (youtube.com)](https://www.youtube.com/watch?v=9GvnrQv138s&list=PLjVLYmrlmjGcC0B_FP3bkJ-JIPkV5GuZR)
* Seaborn: [What is Data Visualization | Python Seaborn Tutorials (youtube.com)](https://www.youtube.com/watch?v=hLbVXF70BCE&list=PLjVLYmrlmjGfhqSO3rF4n02rrj9w2Ch2C)