CVR COLLEGE OF ENGINEERING

(An Autonomous Institution, NAAC Accredited and Affiliated to JNTU, Hyderabad)
Vastunagar, Mangalpalli (V), Ibrahimpatnam (M),Rangareddy (D), Telangana- 501 510

DEPARTMENT OF CSE (DATA SCIENCE)

B.Tech II Year II Semester

Real-Time/Field - Based Research Project Review - 1

Car Price Prediction

By
Artham Ajay(23B81A6706)
K.Pranaydeep(23B81A6727)
M.Sanjana(23B81A6745)

Under the guidance of
Dr.S.V.Suryanarayana Sir
Head of the Department CSE(Data Science)

Domain Introduction

The automotive market is dynamic, with factors like brand, year, mileage, fuel type, and technology advancements significantly influencing vehicle prices.
Accurate car price prediction is essential for buyers, sellers, and dealerships to make informed decisions. Traditional manual estimations are often unreliable and subjective.

By leveraging **Machine Learning (ML)**, especially techniques like regression and ensemble models, we can create data-driven systems that analyze historical trends and vehicle attributes to predict prices more accurately.
This project aims to provide a **fast, efficient, and transparent** method of estimating car values, ultimately improving trust, efficiency, and stability in the used car market.

Challenges:

•Car prices depend on multiple dynamic factors such as brand, model, mileage, fuel type, vehicle age, transmission, etc.

•Capturing complex, non-linear relationships between these features is difficult.

•Traditional manual methods for car valuation are inaccurate and subjective.

•Need for a predictive model that generalizes well to unseen data while maintaining high accuracy

Problem Statement:

Develop a robust predictive model using machine learning, specifically the Random Forest Regressor, to estimate the selling price of used cars based on multiple vehicle attributes. The goal is to create a system that delivers accurate, reliable, and data-driven car price predictions to assist both buyers and sellers, enhancing transparency and efficiency in the automotive market.

Existing Works/Methods And challenges

1. CarDekho Dataset Approach

•Provides detailed car information: brand, model, year, mileage, fuel type, seller type, etc.
•Machine Learning techniques applied:
- Regression Analysis
- Decision Trees
- Ensemble Methods

•Limitations:
- Data quality issues (missing/inaccurate features)
- Cannot always capture sudden market trends
- Limited to used car data

2. CARS24 Platform Approach

•Uses massive transaction datasets including car details, ownership, location, insurance status, etc.
•Advanced models used:
- Random Forests
- XGBoost
- CatBoost

•Additional Innovations:
- **Computer Vision**: Analyzes uploaded car images to detect damages
- **NLP Techniques**: Extracts insights from seller descriptions and comments

•Limitation:
- Black-box nature of complex models reduces interpretability for customers

Proposed Method:

Description:

The proposed system will use machine learning algorithms to analyze historical car sales data and predict prices for used cars. The system will be trained using structured datasets containing various car attributes, and it will be capable of making accurate predictions for unseen data.

List of Modules/Features:

- Data Collection-collecting dataset from Kaggle
- Data preprocessing-finding the missing values and filling them with mean,mode and median of features
- Exploratory Data Analysis (EDA)-user for data visualization
- Model Training-splitting the dataset into training set and testing set,performing Random Forest and predicting the prices
- Model Evaluation- Measure the model's accuracy and performance.
- Prediction Interface-Deploy the trained model so users can predict car prices by entering details.

Technology Stack:

Programming Language:

Python: Used for data preprocessing, model development, and deployment.

Libraries:

Pandas: For loading, cleaning, and analyzing datasets.

Scikit-learn: For implementing the Random Forest model and preprocessing.

Gradio: To create a user-friendly web interface for real-time car price prediction.

Joblib: For saving and loading the trained model and encoders.

Environment:

Jupyter Notebook: Interactive development and testing of machine learning code.

Anaconda: Simplifies package management and environment setup.

Model:

Random Forest Regressor: A powerful ensemble learning method used for accurate car price prediction.

Deployment:

Gradio Web App: Allows users to input car details and instantly get price predictions through an intuitive UI

Module Structure:

1. Data Collection and Preprocessing

- Gathered a dataset with car attributes like brand, mileage, fuel type, etc.
- Handled missing values, encoded categorical features using Label Encoding, and cleaned the data.

2. Feature Engineering

- Selected key features such as brand, vehicle age, mileage, engine size, and fuel type.
- Dropped less important features like "model" to reduce noise and improve model accuracy.

3. Model Training and Testing

- Used Random Forest Regressor for training the model.
- Split the dataset into training (80%) and testing (20%) to validate performance.
- Evaluated using R^2 score and Mean Absolute Error (MAE).

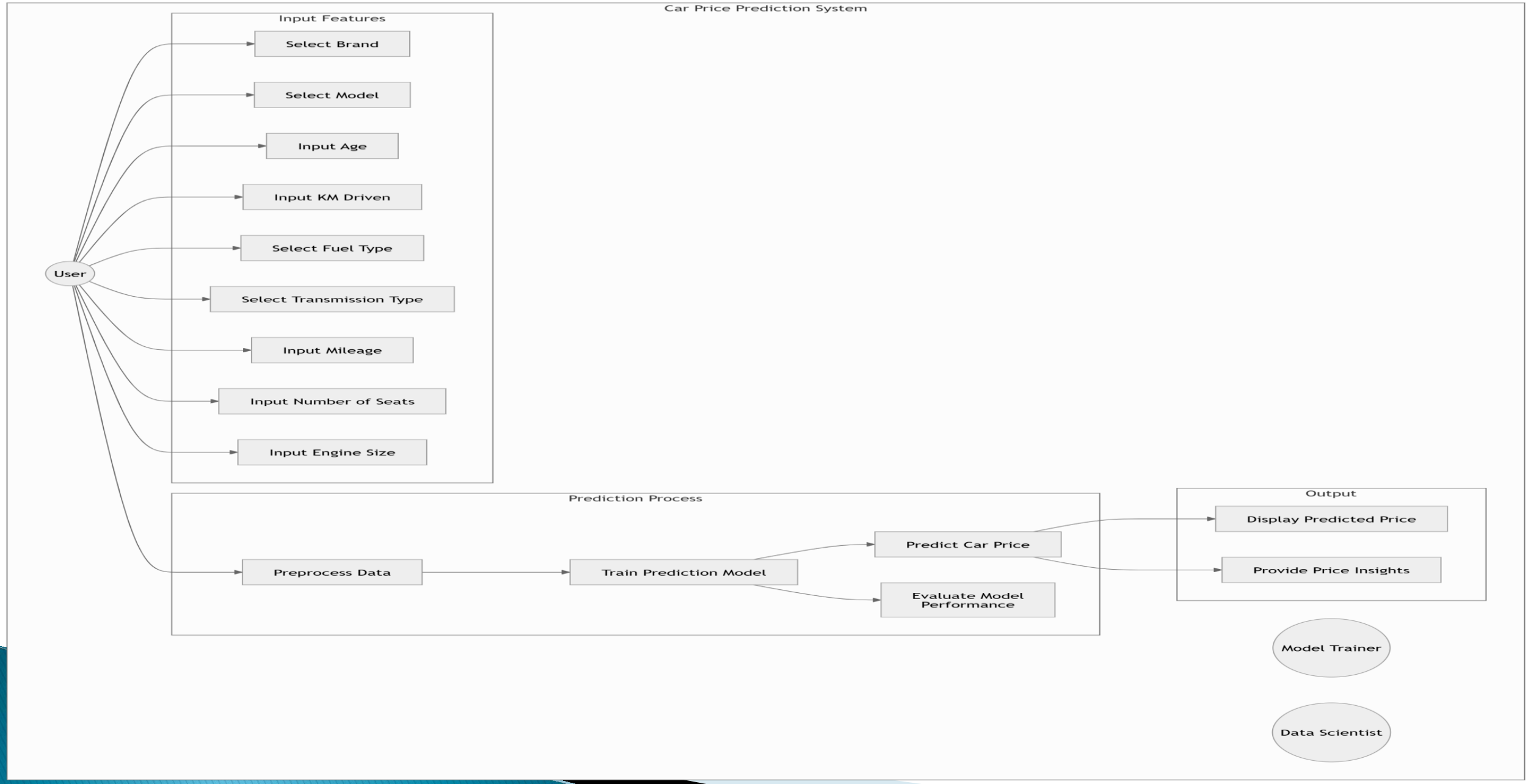
4. Model Saving and Loading

- Saved the trained model and label encoders using **Joblib**.
- Ensured the saved models can be reused without retraining during deployment.

5. Web Interface Deployment (Gradio)

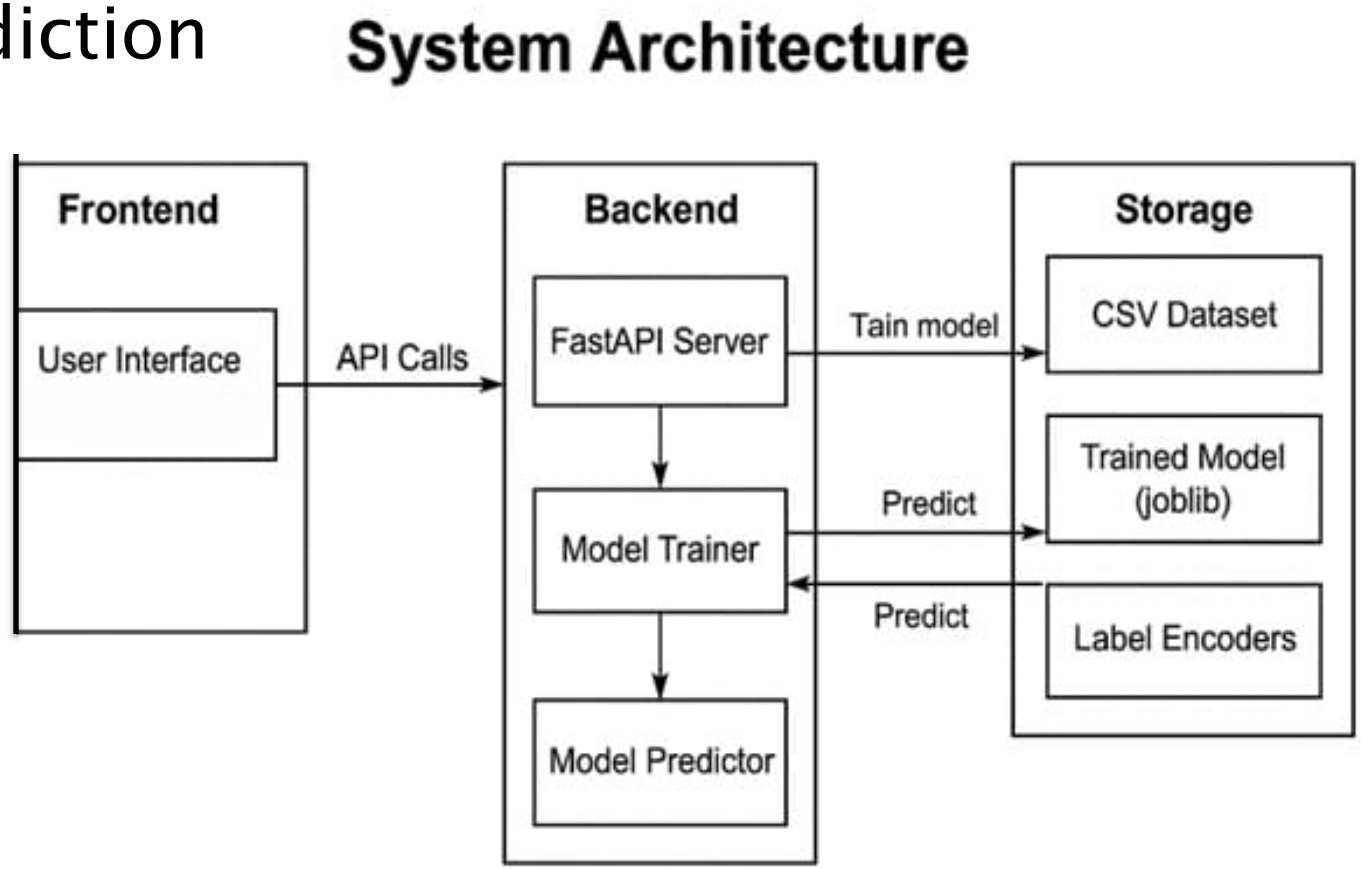
- Built a simple web application using **Gradio**.
- Users can input car details to instantly get a predicted selling price.
- Made the model accessible and user-friendly.

Data Flow Diagram



Architecture diagram:

Car price prediction



Contribution of Students :

Student 1 : A.Ajay

- Contributed Collection ofdataset
- Work on dataset cleaning and preprocessing
- Train the model(Random Forest)
- Integration model files with gradio web app
- Frontend and Backend Connection

Student 2: M.Sanjana

•Contributed in collection of dataset
• Apply Label Encoding on Categorial variables
•Feature Engineering
•Testing the training and testing data finding R^2 Score

Student 3: K.Pranaydeep

- Contributed in collection and cleaning of dataset
- Training and evaluating the model in Multiple Linear regression
- Plotting the relations

Results and Discussion:

Model Used	R^2 score on testing data	R^2 score on training data
Multiple Linear regression	0.85	0.82
Random Forest	0.97	0.90

The car price prediction project demonstrates the effective use of machine learning techniques in estimating the market value of vehicles based on various features such as make, model, year, mileage, fuel type, transmission, and more. By training and evaluating multiple models including Linear Regression and Random Forest the project identifies the most accurate and efficient approach for price estimation.
The results show that ensemble methods like Random Forest outperform simpler models, providing higher accuracy and robustness against overfitting.

Conclusion :

The car price prediction project demonstrates the effective use of machine learning techniques in estimating the market value of vehicles based on various features such as make, model, year, mileage, fuel type, transmission, and more. By training and evaluating multiple models including Linear Regression and Random Forest the project identifies the most accurate and efficient approach for price estimation

Future Scope Real-TimeData Integration:

Integrate live data from car listing websites (e.g., CarDekho, OLX, CarWale) via APIs to ensure dynamic and updated price predictions. Image-BasedPriceEstimation: Use computer vision techniques to estimate vehicle condition from images and factor it into the pricing model

References:

[1] K. Dey, M. A. J. Al Zubair, and M. A. Karim, “Predicting Car Prices using Machine Learning Algorithms,” Proceedings of the 2020 International Conference on Computer Science and Software Engineering (CSSE), 2020, pp. 140-145.

[2] J. R. Hennig, S. G. Hearn, and C. R. Casson, “Car Price Prediction with Regression Models and Neural Networks,” Journal of Artificial Intelligence Research, vol. 59, no. 2, pp. 227-243, 2018.

[3] A. Rahman, R. K. Singh, and S. Mishra, “A Comparative Study of Machine Learning Algorithms for Predicting Car Prices,” Proceedings of the 2019 IEEE International Conference on Data Mining, pp. 244-252, 2019.

[4] M. S. Wazir and A. A. Khushnood, “Predicting Used Car Prices using Regression and Ensemble Methods,” International Journal of Computer Science and Information Security (IJCSIS), vol. 18, no. 5, pp. 64-71, 2020.

THANK YOU