Project Proposal

Car Resale Marketplace Data Mining Project

Project Advisor: Prof. Arindam Ray

Team Name: Beta Busters

Team Members:
1. Amolika Godse
2. Anuraag Gujje
3. Ganesh Talluri
4. Sai Sudheer Vishnumolakala
5. Suraj Shiva Kumar

1. Problem Statement

We propose a data mining project to create a car resale marketplace that enables users to sell and buy resale cars. The key objectives of this project are as follows:

1. Predict the resale price of cars based on their features, helping sellers set competitive prices.
2. Recommend similar cars to potential buyers based on their preferences.

2. Source of the Dataset

We have collected data for this project by scraping car sale listings from the Vroom website using Python libraries such as Selenium and Beautiful Soup. This data includes comprehensive information about various car listings, which will serve as the foundation for our car resale marketplace.

GitHub link of the dataset and the scraping code used to extract it:

https://github.com/surajshivkumar/Used-car-price/blob/main/data/master_data.xlsx
https://github.com/surajshivkumar/Used-car-price/blob/main/src/scrape.py

3. Details of the Dataset

After scraping, data cleaning, and removing unnecessary columns, we have a consolidated final dataset consisting of 2,897 rows and 50 columns. Some of the key columns include:

- Make
- Model
- Year
- Price

- VIN (Vehicle Identification No.)
- Total Miles Covered
- Mileage (MPG)
- Transmission
- Body Style
- Fuel Type
- Engine
- Drive Type

All these features will be crucial for building predictive models for suggesting car prices for potential sellers and creating similar car recommendations for potential buyers.

Our Dataset has 3 types of features primarily. Some of the important ones are listed below

1. **Numerical features** – Miles Covered, Mileage, Tank size, Price
2. **Categorical features** – Make, Model, Body style, Doors, Engine, Transmission, Drive type, Fuel type, Cab style etc.
3. **Binary features** – Bluetooth, Heated Seats, Keyless Ignition, Premium Sound System, Satellite Radio, Sunroof Moonroof, Traction Control, Navigation System, Remote Start, Blind Spot Monitor, Parking Assist System, Adaptive Cruise Control, Full Self-Driving Capability, Rear Seat Entertainment etc.

4. Methodology

To accomplish our project objectives, we outline the following approach:

1. Data Extraction:

- Scraping car sale data from the Vroom website using Selenium and BeautifulSoup.

2. Data Preprocessing:

- Data cleaning: Removing duplicates, handling missing values, and addressing outliers.
- Data scaling: Standardizing or normalizing numerical features for model training.
- Feature selection: Identifying relevant features for predicting car prices and recommendations.

3. Data Exploratory Analysis and Visualization:

- Conduct exploratory data analysis to gain insights into the dataset.
- Visualize the data to understand the distribution and relationships between variables.

4. Building Supervised Learning Models:

- Develop predictive models to estimate car resale prices based on features.
- Utilize regression techniques and model evaluation methods to get the best model fit to our data.

5. Recommendation system Using Unsupervised Learning:

- Apply unsupervised machine learning techniques (e.g., clustering) to group cars based on their characteristics.
- These clusters will be used for recommending similar cars to potential buyers.

6. Building a User-Friendly Interface:

- Integrate the predictive and recommendation models into a user-friendly web interface.
- The interface will provide sellers with price suggestions/predictions and buyers with car recommendations.

4. Project Deliverables

At the end of this data mining project, we aim to deliver:

- A clean and processed dataset ready for modeling.
- Supervised machine learning models for predicting car resale prices.
- Unsupervised machine learning models for car recommendations.
- A user-friendly web interface for users to interact with the marketplace.
- Comprehensive documentation and reports detailing the project methodology and outcomes.

This project will provide a valuable solution for individuals looking to buy or sell cars in the resale market, offering fair pricing and personalized recommendations. It has the potential to transform the car resale industry by making it more efficient and user-friendly.