

Pickl.AI's Extensive Internship Program
INTERNSHIP PROJECT - Developing a Machine Learning
Model to predict the selling price of a used car
Name - Sakshi Bidwai | Batch - D

Project Blog

Introduction:

Welcome to an exciting journey into the world of data-driven predictions and machine learning algorithms. In this blog, we'll explore the development of a powerful machine learning model designed to predict the selling price of used cars. By leveraging historical data and advanced analytics, we aim to empower both buyers and sellers with valuable insights, revolutionizing the way pre-owned vehicles are bought and sold. Let's dive in and discover the fascinating potential of data science in the automotive industry!

My Journey in Pickl.ai:

I have been a part of Pickl.ai's learning programs for quite some time now. Previously, I had taken up a machine learning course from Pickl.ai, which ignited my interest in the Data Science field. In today's data-driven world, Data Science plays a pivotal role, driving advancements across industries, improving products and services, and enhancing overall efficiency and competitiveness. When an opportunity arose for a Data Science internship program at Pickl.ai, it was a no-brainer for me. I knew this would help me to further increase my knowledge in this field and gain valuable and practical insights.

The entire training program was intensive and carefully curated, aiming to make us thorough with the required skills while providing practical hands-on sessions for each topic. The regular live sessions were a nice touch as they created a sense of involvement and helped us understand how to apply the knowledge we gained to industry-related problems.

Throughout the internship, we learned a wide range of skills, including Excel, Python Basics, Object-Oriented Programming, Statistics, Numpy, Pandas, Visualization using Python, Tableau, SQL, Data Exploration, Data Selection, Data Cleaning, and, of course, Machine Learning.

Need for a predict the selling price of a used car:

In today's dynamic and competitive automotive market, accurately determining the fair selling price of a used car is crucial for both buyers and sellers. Traditional methods of pricing, such as relying solely on intuition or using generic pricing guides, often fall short in capturing the nuances of individual vehicles. This is where the need for developing a machine learning model to predict the selling price of a used car arises.

Data Abundance: The automotive industry generates an immense volume of data related to used car sales, encompassing various attributes like make, model, year, mileage, condition, and location. Machine learning algorithms can efficiently process this wealth of information to derive meaningful patterns and correlations.

Personalized Pricing: Each used car possesses unique characteristics that influence its value. A machine learning model can take into account specific details and historical pricing trends to generate personalized and accurate price predictions for individual vehicles.

Informed Decision Making: For prospective buyers, knowing the estimated value of a used car can aid in making informed decisions aligned with their budgets and preferences. On the other hand, sellers can optimize pricing strategies to attract potential buyers without undervaluing their vehicles.

Market Competitiveness: The automotive market is highly competitive, and a well-priced used car can significantly impact its chances of being sold quickly. Machine learning models offer a competitive advantage by enabling sellers to set competitive prices based on real-time market trends.

Efficiency and Scalability: Once trained, a machine learning model can quickly evaluate multiple variables and provide accurate price predictions in a fraction of the time compared to manual calculations. This scalability is essential in the fast-paced world of automotive sales.

Reducing Information Asymmetry: Pricing a used car has traditionally been a challenging task, leading to information asymmetry between buyers and sellers. A machine learning model bridges this gap by providing transparent and unbiased pricing insights.

My Approach for the project:

The first step in developing a Machine Learning Model to predict the selling price of a used car is to gain a comprehensive understanding of the task at hand. This involves delving into the intricacies of the used car market, understanding the various factors that influence car prices, and identifying the challenges in accurately predicting the selling price. Once familiar with the domain, the next step is to obtain and explore the dataset. The dataset should consist of relevant attributes such as car make, model, year, mileage, condition, location, and historical selling prices. Performing basic Exploratory Data Analysis (EDA) is crucial to get insights into the data, understand its structure, and identify potential patterns or correlations.

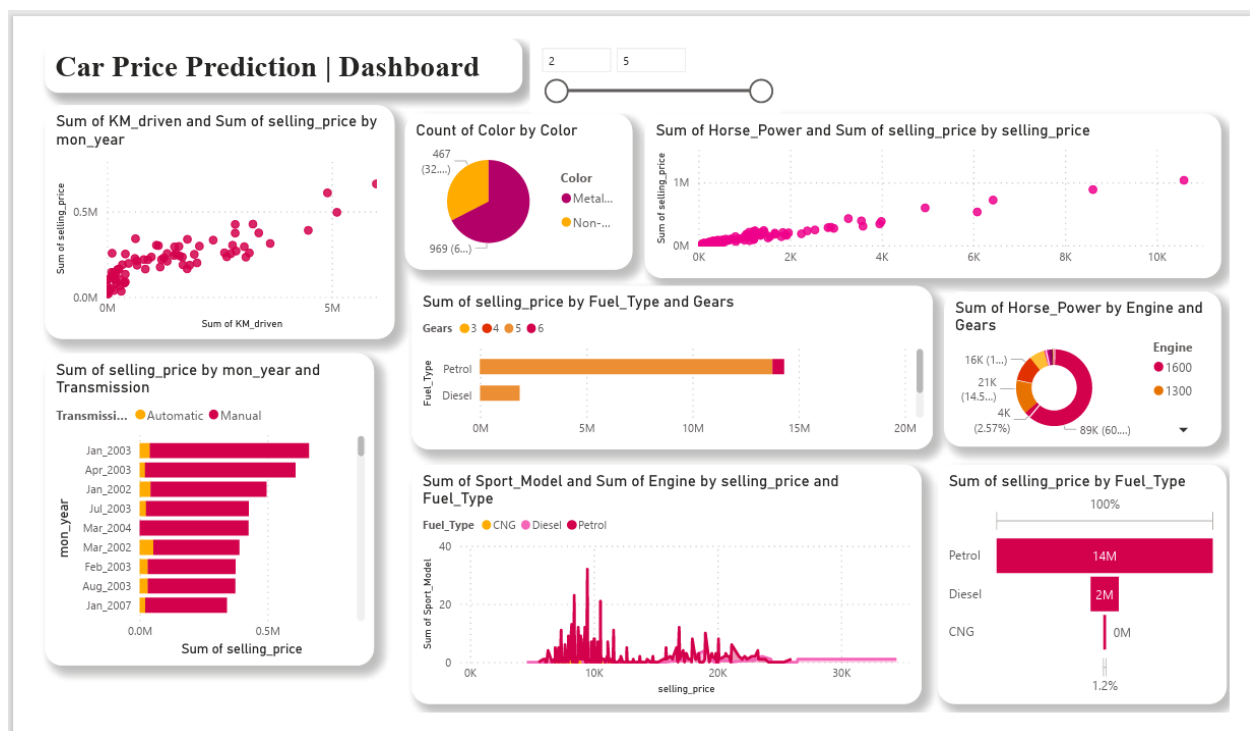
During the EDA process, one may encounter missing values due to the large dataset. To address this, suitable techniques such as mean value imputation can be employed. This involves replacing missing values with the mean of the respective column to ensure the dataset remains consistent and accurate for further analysis. The target variable for our prediction model will be the selling price of the used car. Thus, it becomes essential to focus on understanding the distribution of selling prices, identifying potential outliers, and analyzing how other features correlate with the target variable.

One of the most crucial aspects of the dataset is the 'selling price' column, as it plays a central role in our prediction task. Our end goal is to develop a machine learning model that accurately predicts the selling price of a used car given its relevant features. Depending on the nature of the data and the target variable (continuous or categorical), appropriate machine learning algorithms like regression or classification models can be considered.

Before proceeding to model training, it is imperative to preprocess the data and ensure all features are in a suitable format for analysis. This may involve handling categorical variables using techniques like one-hot encoding, scaling numerical features, and splitting the dataset into training and testing sets for model evaluation.

Throughout the process, continuous refinement and understanding of the data are necessary. By closely examining various features and their distributions, we can identify which factors significantly impact the selling price, thus helping us fine-tune the machine learning model for optimal performance.

Dashboard:



Conclusion:

In summary, developing a Machine Learning Model to predict the selling price of a used car requires a systematic approach that includes understanding the problem, exploring the dataset, handling missing values, conducting EDA, selecting appropriate features, and preparing the data for model training. By following these steps diligently, we can build an accurate and reliable prediction model that will revolutionize the way used cars are priced and traded in the automotive market.

