**Project Proposal: Car Market Trend Prediction & Price Estimation**

**1. Project Overview**

The goal of this project is to develop a robust system that predicts car prices and analyse market trends by leveraging historical car sales data. The system will be powered by machine learning models and feature interactive visualizations, making it an essential tool for car buyers, sellers, and analysts. The key components of this project include:

* A **regression model** to predict car prices based on historical and car feature data.
* An **interactive dashboard** to visualize price trends, sales data, and market trends, developed using Plotly for visualizations.
* A **chatbot** designed to provide car recommendations based on user preferences, powered by the model’s predictions.
* An **automated retraining pipeline** that ensures the model continuously improves as new data is added.

**2. Objectives**

The main objectives of this project are to:

* **Develop a predictive model**: Build a machine learning model using **linear regression** to estimate car prices based on a set of relevant features such as brand, year, and mileage.
* **Perform deep data analysis**: Conduct a thorough analysis of the car sales data to identify key trends, including the most popular car brands, the impact of car condition, and pricing patterns over time.
* **Implement interactive visualizations**: Create dynamic visualizations that evolve as new data is added, allowing users to interact with the data and track market trends.
* **Create a user-friendly interface**: Design a front-end dashboard where users can input car details and receive price predictions, while also exploring sales trends and market data.
* **Build a chatbot**: Develop a conversational interface that suggests cars based on user preferences such as budget, car brand, and desired features.
* **Automate retraining**: Set up an automated retraining pipeline where the model is updated periodically as new, verified data is introduced into the system.

**3. Data Sources**

The dataset used for this project is sourced from Kaggle and contains detailed information about car sales in Australia:

* **Dataset**: [Australia Car Market Data](https://www.kaggle.com/datasets/lainguyn123/australia-car-market-data)
* **Features Used**:
  + **Price**: The target variable for price prediction.
  + **Brand & Model**: Categorical features representing the car brand and model.
  + **Year**: Year of manufacture.
  + **Kilometres Driven**: A continuous feature indicating the distance driven by the car.
  + **Fuel Type**: Categorical feature for the type of fuel used (e.g., Petrol, Diesel, Electric).
  + **Seating Capacity**: Numeric value representing the number of seats.
  + **Gearbox Type**: Categorical feature (Automatic/Manual).
  + **Car Condition**: A categorical feature indicating whether the car is new or used.

**4. Machine Learning Approach**

* **Regression Model**:
  + **Algorithm**: Linear Regression (using Scikit-learn)
  + **Target Variable**: Price (the price of the car)
  + **Feature Engineering**: The dataset undergoes preprocessing, including one-hot encoding for categorical variables, handling missing values, and standardizing numerical features for improved model performance.
* **Active Learning & Retraining**:
  + The model will actively improve as new data becomes available, with the performance continually refined through human verification before retraining.
  + **Retraining Process**: The model will retrain automatically once a specified number of new data points have been verified by an administrator, ensuring it remains up-to-date with evolving trends.

**5. Technologies & Tools**

The following tools and frameworks will be used to implement this project:

| **Category** | **Tools/Frameworks** |
| --- | --- |
| **Machine Learning & Data Processing** | Scikit-learn, Pandas, NumPy |
| **Data Visualization** | Matplotlib, Plotly |
| **Dashboard & Front-end** | Flask (API), React.js (UI) |

**6. Expected Deliverables**

The final product will include the following key deliverables:

1. **Car Price Prediction Model**: A regression-based model capable of accurately predicting car prices based on various features.
2. **Interactive Dashboard**: A dashboard built with Plotly that visualizes car market trends, including heatmaps, bar charts, scatter plots, and other visual components to track key market indicators.
3. **Car Recommendation Chatbot**: An AI-powered chatbot that offers car recommendations based on user inputs such as price range, brand preferences, and car features.
4. **Market Trend Visualization**: Visualizations to identify clusters of car sales trends using machine learning, helping users understand regional variations, pricing patterns, and feature demand.
5. **Automated Model Retraining**: An automated pipeline that continuously retrains the model as new data is verified, ensuring its predictions stay relevant and accurate.

**7. Impact and Conclusion**

This project will provide car buyers, sellers, and market analysts with a powerful tool to estimate car prices, analyze market trends, and receive personalized car recommendations. The combination of machine learning, interactive visualizations, and automated model retraining will significantly enhance decision-making and improve market insights. By continuously adapting to new data, the system will offer up-to-date predictions and trends, making it a valuable resource for the automotive industry.