**TOPIC: Car Price Prediction Using Machine Learning**

* **Problem Definition:**

Predicting car prices accurately based on various factors like make, model, year, mileage, and condition is a critical problem in the automotive sector. This task benefits both buyers and sellers in making informed decisions. It also enables manufacturers to refine pricing strategies and forecast market trends.

**Significance:**

* Facilitates informed purchasing and selling decisions.
* Optimizes manufacturing and marketing strategies in the automotive industry.
* Contributes to research in predictive analytics within the automobile sector.

**Objectives:**

* Develop a machine learning model for accurate car price predictions.
* Compare and evaluate the performance of multiple machine learning algorithms.
* **Asking the Right Questions:**

What are the key factors influencing car prices?

How can we preprocess data effectively to improve model accuracy?

Which machine learning algorithm provides the best performance for this dataset?

Can external factors like market trends or location data improve model predictions?

* **Data Collection:**

The dataset for this project was collected from a source containing 301 records, including attributes such as:

* Kilometers traveled.
* Year of registration.
* Fuel type.
* Car model.
* Fiscal power.
* Car brand.
* Gear type.
* **Data Wrangling (Preprocessing):**

Steps undertaken:

1. **Handling Missing Values:** Removed missing or irrelevant data.
2. **Feature Engineering:** Added the "Age of Car" column by subtracting the year of registration from the current year and removed the redundant "Year" column.
3. **Outlier Detection:** Identified and addressed outliers using visual graphs.
4. **Data Standardization:** Standardized numerical variables to ensure consistent scaling.

* **Exploratory Data Analysis (EDA):**

Performed EDA to identify relationships and patterns in the dataset:

* Visualized correlations between price and other factors such as mileage, age, and fuel type.
* Highlighted that price negatively correlates with the car's age and kilometers traveled.
* Tools: Python libraries such as pandas, matplotlib, and seaborn were used.
* **Predictive Analysis:**

Machine learning models applied:

**Linear Regression:**

Achieved 70% accuracy.

Models linear relationships between price and car features.

**Random Forest Regressor:**

Accuracy: 75%.

Ensemble learning technique aggregating decision trees.

**Gradient Boosting Regression:**

Accuracy: 87%.

Boosting technique offering improved accuracy.

**Extreme Gradient Boosting Regression (XGB):**

Accuracy: 90% approx.

Outperformed other models with the lowest MSE and highest robustness.

Evaluation Metrics:

R², RMSE, and accuracy were used to assess model performance.

* **Deliverables:**

Model Evaluation: Gradient Boosting and XGBoost Regression demonstrated the best results.