A logo of a company

Description automatically generated A close-up of a logo

Description automatically generated

**Joint Tech Internship Community Program**

**Assignment: Predicting Car Resale Prices in the Automotive Market**

**Problem Statement:**

An online car dealership wants to predict the resale prices of used cars based on various features such as the car's make, model, age, mileage, and other relevant characteristics. Your task is to develop a machine learning model that predicts the resale price of a car given its attributes.

**Objective:**

Build a predictive model to estimate the resale price of used cars. Evaluate the model using appropriate metrics and provide insights into the factors that most influence car resale prices.

**Dataset:**

You are provided with a dataset containing the following columns:

1. **CarID**: Unique identifier for each car.
2. **Make**: The manufacturer of the car (e.g., Toyota, Ford, BMW).
3. **Model**: The specific model of the car (e.g., Corolla, Focus, 3 Series).
4. **Year**: The year the car was manufactured.
5. **Mileage**: The total miles the car has been driven.
6. **EngineSize**: The size of the car's engine (in liters).
7. **FuelType**: The type of fuel the car uses (e.g., Petrol, Diesel, Electric, Hybrid).
8. **Transmission**: The type of transmission (e.g., Manual, Automatic).
9. **Color**: The color of the car.
10. **PreviousOwners**: The number of previous owners the car has had.
11. **CarCondition**: The overall condition of the car (e.g., Excellent, Good, Fair, Poor).
12. **Location**: The location where the car is being sold (e.g., Urban, Suburban, Rural).
13. **ResalePrice**: The target variable representing the estimated resale price of the car (in USD).

**Tasks:**

1. **Data Exploration and Preprocessing:**
   * Load the dataset and perform initial exploration to understand the data.
   * Identify and handle any missing values appropriately.
   * Convert categorical variables into numerical ones using techniques such as One-Hot Encoding.
2. **Feature Engineering:**
   * Perform feature scaling (e.g., Standardization or Normalization) on continuous variables.
   * Create new features if relevant, such as age of the car at the time of sale.
   * Use dimensionality reduction techniques (e.g., PCA) to reduce the feature space if necessary.
3. **Model Building:**
   * Split the dataset into training and testing sets (e.g., 80-20 split).
   * Train different regression models (e.g., Linear Regression, Decision Trees, Random Forest, Gradient Boosting).
   * Perform hyperparameter tuning using techniques like Grid Search or Random Search to optimize model performance.
4. **Model Evaluation:**
   * Evaluate your models using appropriate metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.
   * Compare the performance of different models and select the best one.
   * Analyze feature importance to understand the most significant factors contributing to car resale prices.
5. **Insights and Recommendations:**
   * Provide insights based on your model analysis regarding the factors that influence car resale prices.
   * Suggest actionable strategies for car dealerships to optimize their pricing strategies based on the findings.
6. **Documentation:**
   * Document your process, including data exploration, preprocessing steps, model selection, and evaluation.
   * Include visualizations where necessary to support your findings.

**Deliverables:**

* A Jupyter notebook (or Python script) with the entire workflow.
* A report summarizing your findings, including the model's performance and recommendations for optimizing car resale prices.

This assignment problem focuses on predicting car resale prices, a linear regression problem in the automotive domain. The problem involves data preprocessing, feature engineering, model training, and evaluation, with a focus on understanding the factors that drive car resale values.