# IMARTICUS LEARNING  BANGALORE

# Title: Bike Resale Price Prediction Using Machine Learning

## Subtitle: A Detailed Explanation of Data Analysis and Model Building

BY :SHIVANAND M A

# Overview of the project

- **State –** The geographical region or state where the bike is being sold. This helps analyze trends in bike sales across different regions.
- **Avg Daily Distance (km)** – The average distance traveled by the bike per day. It can indicate how much the bike has been used, affecting its resale price.
- **Brand** – The manufacturer of the bike (e.g., Honda, Yamaha, Royal Enfield). Brand reputation can impact pricing and demand.
- **Model** – The specific model name of the bike (e.g., Pulsar 150, Splendor Plus). Different models have varying resale values based on features and popularity.
- **Price (INR)** – The original price of the bike when purchased new. This helps in comparing the depreciation over time.

- **Year of Manufacture** – The year the bike was manufactured. Older bikes tend to have lower resale values.
- **Engine Capacity (cc)** – The displacement of the bike's engine in cubic centimeters (cc). Higher cc bikes generally have more power but may have lower mileage.
- **Fuel Type –** The type of fuel the bike uses (e.g., Petrol, Electric). Fuel efficiency and running costs impact resale value.
- **Mileage (km/l)** – The fuel efficiency of the bike, measured in kilometers per liter (km/l). Higher mileage bikes are often more desirable.
- **Owner Type** – Indicates whether the bike is being sold by the first owner, second owner, or beyond. First-owner bikes generally have higher resale prices.

- **Registration Year** – The year the bike was registered for use on the road. This may differ from the year of manufacture and affect legal and resale aspects.
- **Insurance Status** – Indicates whether the bike has active insurance. Insured bikes are generally more valuable.
- **Seller Type** – The type of seller (e.g., Individual, Dealer). Dealers may offer slightly higher prices due to warranties or service history.
- **Resale Price (INR)** – The estimated price at which the bike can be resold. This is the target variable for the prediction model.
- **City Tier** – Categorization of the city where the bike is being sold (e.g., Tier 1, Tier 2, Tier 3). Larger cities (Tier 1) may have higher demand and resale prices

# Problem Statement: Bike Resale Price Prediction Using Machine Learning

The second-hand bike market is growing rapidly, with numerous factors influencing resale prices. Buyers and sellers often struggle to estimate the fair market price of a used bike, leading to inaccurate pricing, inefficient transactions, and financial losses. A data-driven machine learning model can help predict resale prices based on various bike attributes, ensuring fairness and transparency.

- Accurate Price Estimation – A well-trained model that predicts fair resale values for bikes.
- Market Insights – Understanding how different factors influence pricing trends.
- Better Decision-Making – Buyers and sellers can use the model to negotiate fair deals.
- Scalability – Potential integration with e-commerce platforms like OLX, BikeDekho, and Droom.

# Exploratory Data Analysis(EDA)

## 1. **Understanding the Dataset**

- Load the dataset using Pandas (pd.read_csv()).
- Display the first few rows using df.head().
- Check the structure using df.info(), which shows column names, data types, and non-null values.
- Identify missing values using df.isnull().sum().

## 2.Descriptive Statistics

- Use df.describe() to get statistical summaries (mean, median, standard deviation, min/max values).
- Analyze numeric columns like Price, Mileage, Engine Capacity, Resale Price for distributions.
- raph text

## 3.Data Visualization

To identify patterns and relationships:

- Histogram: sns.histplot(df['Price']) (Visualizes price distribution)
- Boxplot: sns.boxplot(x='Fuel Type', y='Price', data=df) (Shows price variations by fuel type)
- Pairplot: sns.pairplot(df[['Price', 'Mileage', 'Engine Capacity']]) (Shows feature relationships)
- Correlation Heatmap: sns.heatmap(df.corr(), annot=True, cmap='coolwarm') (Shows feature correlations)

### 4.Feature Engineering

- Convert categorical features (Brand, Model, Fuel Type, Seller Type) using one-hot encoding or label encoding.
- Create new meaningful features (e.g., Age of the Bike = Current Year - Year of Manufacture).
- Scale numerical features using StandardScaler or MinMaxScaler.

## 5. Outlier Detection & Removal

- Use boxplots and z-score method to identify outliers.
- Remove extreme values in Price, Mileage, and Resale Price.

## 6. Splitting Data for Model Building

- Define features (X) and target variable (y).
- Split dataset into training (80%) and testing (20%) using train_test_split().

# Machine Learning Models Consideredt

## (a) Linear Regression

✅ Best for simple relationships between independent variables and the target variable.

❌ Does not work well when data has non-linearity or complex interactions.

**When to Use:**

- When the data has a linear trend
- When the dataset is small and well-structured

## b) Decision Tree Regressor

✅ Works well for non-linear relationships.

✅ Captures feature interactions automatically.

❌ Can overfit if not pruned properly.

aph text

Back to Agenda

**When to Use:**

- When the data has non-linear relationships
- When we need rule-based decision-making

## (c) Random Forest Regressor

✅ Best for high accuracy and generalization.

✅ Reduces overfitting by averaging multiple decision trees.

❌ Computationally expensive (slower training time).t

**When to Use:**

- When we need high accuracy and robust predictions
- When the dataset is large and complex

Back to Agenda

## Best Model for Bike Resale Prediction

✅ Random Forest Regressor is the best model because it provides the lowest error and highest accuracy.

✅ It captures complex interactions between features without overfitting like a single Decision Tree.

✅ However, if computational power is a concern, Decision Tree Regressor can also be considered.

# Future Scope of Machine Learning in Bike Resale Price Prediction

## 1. Integration with Real-Time Market Trends

- Web Scraping: Collect live resale price data from bike-selling platforms (e.g., OLX, BikeDekho, Droom).
- API Integration: Use APIs from automotive marketplaces to update bike prices dynamically.
- Macroeconomic Factors: Incorporate fuel price trends, inflation rates, and demand-supply factors for better predictions.

## 2. Expanding Features for Better Predictions

- Sentiment Analysis: Extract customer reviews and ratings to assess bike demand.
- GPS & IoT Data: Use real-time mileage and maintenance data from connected vehicles.
- Vehicle Condition Assessment: Integrate bike condition reports (accident history, servicing details).

## 3. Deploying the Model for Public Use

- Web Application: Create a user-friendly interface where users can enter bike details and get a predicted resale value.
- Mobile App Integration: Develop a mobile app for on-the-go predictions.
- Cloud Deployment: Host the model on cloud platforms like AWS, Google Cloud, or Azure for scalability.

## 4. Future Research Directions

- Multi-Modal Learning – Combine text, images, and structured data to enhance predictions.
- Reinforcement Learning – Optimize pricing strategies dynamically based on demand patterns.
- Explainable AI (XAI) – Develop interpretable models to provide reasons for each predicted price.

ext

# Conclusion of Bike Resale Price Prediction using Machine Learning

In this project, we used machine learning techniques to analyze and predict the resale price of bikes based on various features like brand, model, engine capacity, mileage, fuel type, owner type, and city tier. We followed a structured approach:

✅ The project successfully developed a machine learning model to predict bike resale prices with high accuracy.

✅ Random Forest Regressor emerged as the best-performing model.

✅ Further improvements can be made by integrating real-time data and advanced AI techniques.

✅ This model has practical applications for sellers, buyers, and online marketplaces.text

# THANK YOU