House Price Prediction Using Machine Learning

# 1. Objective:

The aim of this project is to predict house prices using different machine learning models based on various house-related features like Area, Bedrooms, Bathrooms, Garage, Location, and Condition.

# 2. Dataset Description:

• Total Records: 2,000  
• Columns: Id, Area, Bedrooms, Bathrooms, Floors, YearBuilt, Location, Condition, Garage, Price  
• No missing values were found.  
• 3 categorical features: Location, Condition, Garage.

# 3. Tools & Libraries Used:

• Python, Google Colab  
• Pandas, NumPy, Matplotlib, Seaborn  
• scikit-learn, XGBoost

# 4. Steps Performed:

1. Imported the dataset into Google Colab using Pandas.  
2. Explored the data using .head(), .info(), .isnull().sum().  
3. Created new features: HouseAge = 2025 – YearBuilt, TotalRooms = Bedrooms + Bathrooms.  
4. Encoded categorical columns (Location, Condition, Garage) using One-Hot Encoding.  
5. Dropped unnecessary columns like Id, YearBuilt, Bedrooms, Bathrooms (after combining).  
6. Visualized data with scatter plots, boxplots, and heatmaps to check feature relationships.  
7. Split data into training and testing sets (80:20).  
8. Trained multiple models: Linear Regression, Decision Tree Regressor, Random Forest Regressor, XGBoost Regressor.  
9. Evaluated models using MAE, MSE, RMSE, and R² score.  
10. Compared feature importance to check which features impacted the price the most.

# 5. Results:

• Linear Regression: Negative R² score.  
• Decision Tree: Negative R² score.  
• Random Forest: Negative R² score.  
• XGBoost: Negative R² score.  
All models gave low or negative R² scores because the dataset lacked strong predictors for house price.

# 6. Conclusion:

The models could not achieve high accuracy due to limited features. Real-life house price prediction requires more detailed data, such as exact location (latitude/longitude), neighborhood amenities, property condition, nearby facilities, and market trends.

# 7. Future Suggestions:

• Collect more relevant data such as crime rate, proximity to schools and hospitals, market value trends, and exact location coordinates.  
• Try more advanced models with hyperparameter tuning.  
• Visualize more advanced correlations for deeper insights.

Prepared By:  
Ebad Khan  
Learning  
Institute Name: Emerging India Institute