# Problem Statement

# Problem Statement:

# Accurately forecasting house prices is a key challenge in real estate due to the influence of multiple dynamic factors like location, size, and market trends.

# Importance:

# 1. It helps buyers and sellers make informed, fair decisions.

# 2. Banks use it to assess mortgage and investment risks.

# 3. Developers plan better with price trend insights.

# 4. Governments use forecasts for effective housing policies.

# 5. It enables application of advanced regression techniques in real-world scenarios.

# Objectives of the Project

* + The primary objective of this project is to develop a machine learning model that can accurately predict the price of a house based on various features such as location, size, number of bedrooms, and other important factors.
  + The model uses smart regression techniques, which may involve a combination of linear regression, advanced ensemble methods like XGBoost or LightGBM, and extensive feature engineering.

# Scope of the Project

* Real-time data integration from live property websites
* Web/mobile app deployment for user-friendly access
* Incorporating more features like proximity to schools, hospitals, and public transport
* Multi-city support with dynamic datasets
* Deep learning integration for improving accuracy and learning complex patterns
* Integration with GIS systems for map-based predictions

# Data Sources

Common Data Sources for House Price Forecasting:

1. Kaggle Datasets – Well-known datasets like the House Prices: Advanced Regression Techniques dataset.

# High-Level Methodology

# High-Level Methodology for House Price Forecasting Model:

# 1. Data Collection

# Gather data from reliable sources such as Kaggle, Zillow, or government property records.

# 2. Data Preprocessing

# Clean the data by handling missing values, encoding categorical features, and normalizing numerical data.

# 3. Feature Engineering

# Create new meaningful features (e.g., age of house, price per sq. ft.), and select the most relevant ones.

# 4. Model Selection and Training

# Use smart regression techniques like Linear Regression, Ridge, Lasso, or Gradient Boosting to train the model.

# 5. Model Evaluation and Tuning

# Evaluate using metrics like RMSE or R², and tune hyperparameters for better accuracy using cross-validation.

# 6. Prediction and Deployment

# Use the final model to predict house prices on unseen data and deploy it for real-time usage if needed.

# Tools and Technologies

**Hardware Requirements:**

A Laptop with 8GB RAM

Intel i3 or equivalent processor

**Software Requirements:**

* Python 3.x
* VS Code

Required Python Libraries:

* pandas
* NUMPY
* matplotlib
* seaborn
* scikit-learn
* XGBOOST
* LIGHTGBM
* StreamLit

**OPERATING SYSTEM:**

Windows 11