**Introduction:** The housing market is big. Predicting house prices helps buyers, sellers and investors make good decisions. Data science lets us look at past trends and improve accuracy. This project uses machine learning on real estate data to find out what matters and give insights.

**Problem Statement:** This project will create a model to predict house prices. The goal is to find out what matters and improve the accuracy using machine learning. By looking at historical data, the project will help investors and businesses make better choices.

**Data Collection and Preprocessing:** We used publicly available real estate data. Before using it, we cleaned and prepared the data by:

* Handling missing values (LotFrontage, MasVnrType, GarageType).
* Converting categorical data into numerical form using label encoding and one-hot encoding.
* Scaling numerical features to make sure everything is on the same scale.

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

* Strong correlation between features and house prices.
* Trends of how location, size and facilities affect prices.
* Outliers that can affect model performance.

**Feature Engineering:** We created new features by:

* Combining existing features to capture complex relationships.
* Applying transformations to simplify data patterns.
* Adding neighborhood average prices to give context.

**Model Selection and Training:** We tried multiple machine learning models:

* **Linear Regression:** A simple model for baseline.
* **Decision Trees and Random Forest:** Capture complex relationships well.
* **Gradient Boosting (XGBoost):** High accuracy.

**Hyperparameter Tuning:** We tuned the models using grid search and random search to find the best settings and prevent overfitting.

**Model Evaluation:** We evaluated the models using:

* **Mean Squared Error (MSE):** Measures the accuracy.
* **R-squared (R2):** How well the model explains the price variations.
* **Root Mean Squared Error (RMSE):** An understandable error measure.

Results:

* **Linear Regression:** RMSE = 35,460, R2 = 0.853
* **Decision Tree:** RMSE = 37,210, R2 = 0.821
* **Random Forest:** RMSE = 28,940, R2 = 0.892
* **Gradient Boosting:** RMSE = 27,860, R2 = 0.902**Feature Importance:** What matters:
* **House size** (square feet and number of rooms).
* **Location** (neighborhood and surroundings).
* **Nearby facilities** (schools, transport, parks).

**Business Impact:** Our model benefits **Surprise Housing** by:

* Finding profitable investment opportunities.
* Understanding market trends for better pricing.
* Reduce risks with data-driven insights.

**Conclusion and Next Steps:** We have created a model that predicts house prices. Next steps:

* Add more data (economic indicators).
* Use explainable AI to improve results.
* Try deep learning models for higher accuracy.

Our model helps businesses make better real estate decisions and make more money.