**Predicting House Prices Using Random Forest Regression**

**Introduction**

* **Objective:**
  + The goal of this project is to develop a Random Forest Regression model to predict house prices using the Ames Housing Dataset.
  + This dataset contains various features describing properties, including lot size, number of bedrooms, overall condition, and neighborhood.
  + The project involves data preprocessing, feature selection, model training, evaluation, and visualization of results to ensure accurate price predictions.
* **Dataset Used:** Ames Housing dataset

**Dataset Link:**

[**https://www.openintro.org/data/csv/ames.csv**](https://www.openintro.org/data/csv/ames.csv)

* **Approach:**
  + Data preprocessing, including handling missing values and feature engineering.
  + Split the data into Training and Testing Sets.
  + Training a Random Forest regression model.
  + Evaluating model performance using appropriate metrics.

**Data Preprocessing**

* Handled missing values by dropping irrelevant data to avoid excessive data loss.
* Missing values in categorical and numerical features are replaced with Most Frequent and Median value with respect to columns.
* Machine learning models require numerical input, categorical variables are encoded using Label Encoding.
* Defining feature matrix and target variables.

**Model Selection & Training**

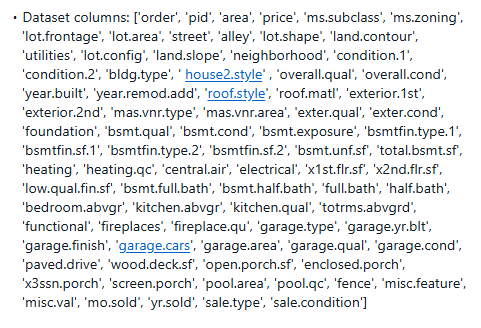
* Chose Random Forest Regression due to its robustness and ability to capture non-linear relationships.
* Split the data into training and testing sets (e.g., 80-20 split).
* Tuned hyperparameters using GridSearchCV (parameters such as the number of estimators, max depth, and min samples split).
* Trained the random forest model on the processed dataset.

**Model Evaluation**

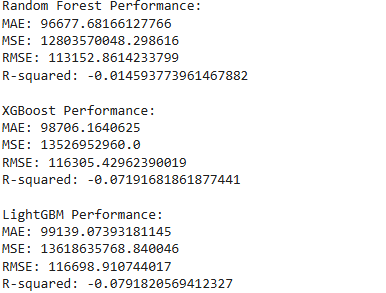
* **Metrics Used:**
  + Mean Squared Error (MSE)
  + Root Mean Squared Error (RMSE)
  + Mean Absolute Error (MAE)
  + R-squared (R²)
* **Performance Analysis:**
  + Evaluated the model on the test set.
  + Compared predicted vs actual prices.
  + Assessed feature importance to understand key factors influencing price predictions.
* **Feature Importance:**
  + Overall Quality - A key determinant of house value.
  + Basement Area - Larger basements often contribute to higher prices.
  + Living Area (Above Ground) - The main living space size significantly impacts pricing.
  + Garage Area - A larger garage adds value to a house.
  + Number of Bathrooms - More bathrooms generally increasehouse value**.**

**Results & Visualizations**

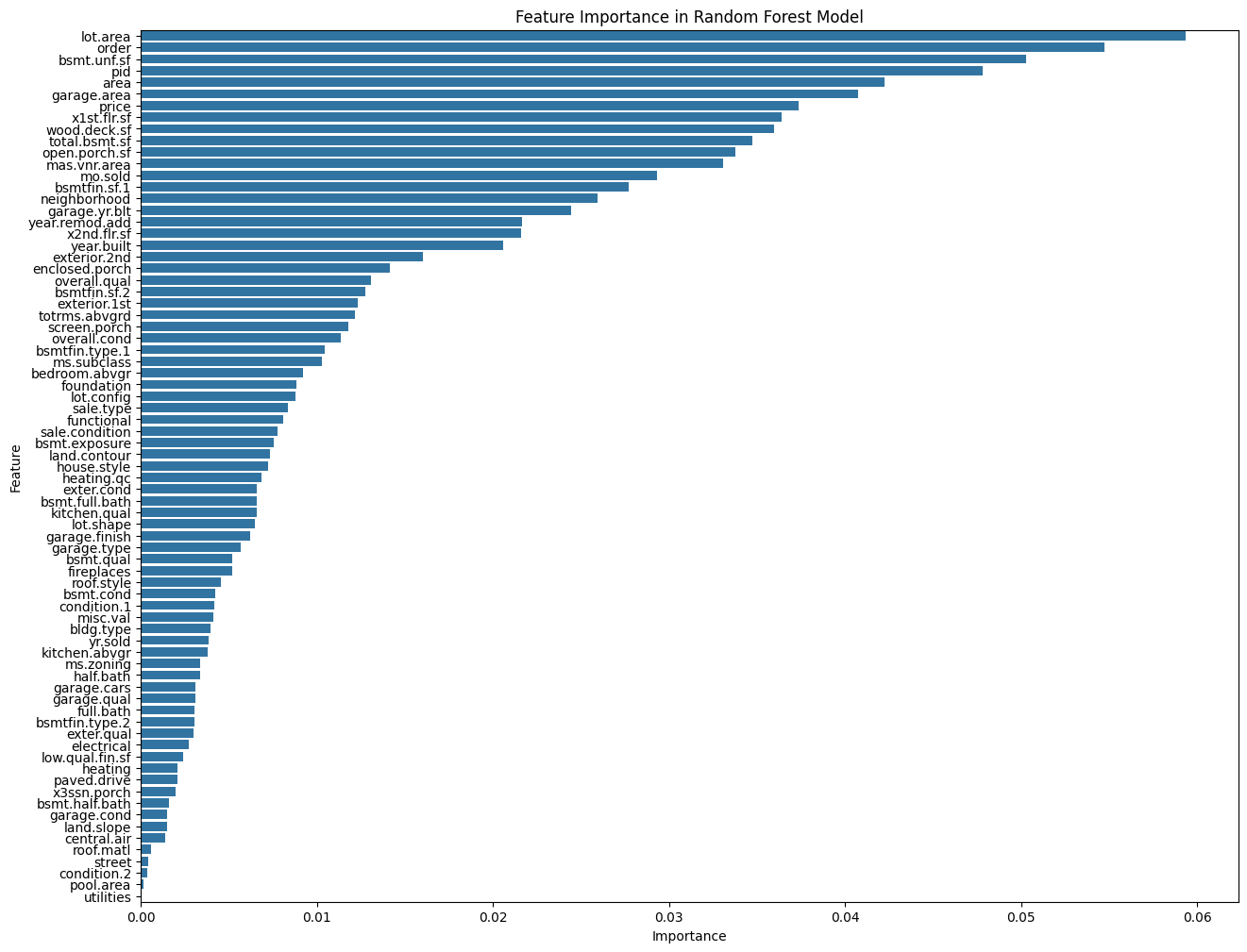
* Feature importance plot highlighting the most influential variables.
* RMSE and R² scores to measure model effectiveness.

Dataset Columns:

Model Evaluation:



Feature Importance:



**Conclusion & Next Steps**

* + **Summary of Findings:**
    - Random Forest provided a reliable model for predicting house prices with reasonable accuracy.
    - Feature importance analysis indicated key drivers such as lot area, overall quality, and living area.
  + **Future Improvements:**
    - Experimenting with ensemble models like XGBoost for improved accuracy.
    - Fine-tuning feature engineering techniques.
    - Using advanced techniques such as stacking or boosting for better generalization.

**Links**

Notebook Link:

<https://colab.research.google.com/drive/1BB5kAfTf7JInPxbL-BvCnsjP0zPbpxzL?usp=drive_link>

Drive Link which includes Dataset, Notebook:

<https://drive.google.com/drive/folders/1j3i3hHZNRt2gkGmPeBSbQFfDgsvngH_0?usp=sharing>

Preferred Platform (Used Platform to deploy this Assignment):

Google Colab: <https://colab.research.google.com>