Problem Statement Worksheet (Hypothesis Formation)

Despite the availability of vast data on residential properties, accurately predicting house prices remains challenging due to the complex interplay of various factors.

H

1 Context

The real estate market is characterized by a high degree of variability in property prices due to factors such as location, size, amenities, and market trends. Accurately predicting house prices is crucial for stakeholders including real estate agents, homeowners, and potential buyers. However, existing methods often struggle to capture the complexity of these factors, leading to suboptimal pricing decisions.

2 Criteria for success

Achieve a significant improvement in the accuracy of house price predictions compared to existing methods.
Develop a machine-learning model that can reliably predict house prices based on relevant features.
Ensure the model's predictions provide valuable insights for stakeholders in the real estate market.
Deploy the model as a user-friendly web application or API for real-time predictions.

3 Scope of solution space

Utilizing the House Prices: Advanced Regression Techniques dataset from Kaggle, which contains a comprehensive set of features related to residential properties.
Performing data preprocessing, including cleaning, missing value imputation, and feature engineering to prepare the dataset for model training.
Experimenting with various machine-learning algorithms such as linear regression, decision trees, random forests, gradient boosting, and neural networks to find the most suitable model.
Evaluating the performance of the trained models using appropriate evaluation metrics such as mean absolute error, mean squared error, and R-squared.

4 Constraints within solution space

Limited availability of data beyond what is provided in the Kaggle dataset may restrict the model's ability to capture all relevant factors influencing house prices.
The model's predictions should be interpretable and explainable to instill trust and confidence among users.

5 Stakeholders to provide key insight

- Real estate agents
- Homeowners
- Potential buyers
- Data scientists
- Software developers

6 Key data sources

- House Prices: Advanced Regression Techniques dataset from Kaggle
- Additional data sources may include publicly available property listings, demographic data, and economic indicators for further analysis.

H D E I P