PREDICTING HOUSE PRICES USING MACHINE LEARNING

APPROACH

Our plan in developing this machine learning model to predict house prices involves all the technology mentioned below along with the machine learning algorithms. The machine learning algorithms that we've decided upon will help us build this project. While developing the model, we might encounter some challenges that would make us add more features/algorithms to enhance the performance of the model. But for now, the below mentioned Algorithms and techstacks will be sufficient for us in making a functional model.

TECH STACK

Programming Languages and Libraries

Python: A versatile and widely-used language for data science and machine learning. Libraries: NumPy, pandas, scikit-learn, TensorFlow, Keras, PyTorch for data manipulation, machine learning, and deep learning.

Data Handling and Analysis

Jupyter Notebook: An interactive environment for data exploration and experimentation.
Data Visualization: Matplotlib, Seaborn, or Plotly for visualizing data and model performance.

Data Preprocessing

Pandas: For data cleaning, feature engineering, and data transformation.
Scikit-learn: For handling missing values, scaling features, and encoding categorical variables.

MACHINE LEARNING ALGORITHMS AND TECHNIQUES TO USE:

Linear Regression: A fundamental algorithm for regression tasks, which establishes a linear relationship between features and target variables.

Gradient Boosting (Optional) : Algorithms like XGBoost, LightGBM, and CatBoost for ensemble learning, which often deliver great results in regression tasks.

Feature Engineering: Create new features, handle missing data, encode categorical variables, and perform scaling to enhance model performance.

Natural Language Processing (NLP) (Optional): Employ NLP methods to extract valuable information from text data such as property descriptions, reviews, and comments.

Ensemble Techniques: Combine multiple models using techniques like stacking, bagging, or boosting to improve predictive accuracy and reduce model variance.

Decision Trees and Random Forests: Effective for capturing non-linear relationships and feature importance. Random Forests offer robust results.