PREDICTING HOUSE PRICES USING MACHINE LEARNING

PHASE 4:Development part 2

Introduction:

In the pursuit of accurate house price predictions, feature selection, model training, and evaluation stand as pivotal stages. Feature selection trims away noise, while model training harnesses algorithms for insights. Evaluation metrics like MAE, MSE, RMSE, and R2 are crucial gauges of success.

Feature Selection:

Feature selection is a critical step to improve the model's performance and reduce complexity. Here are some methods you can consider:

1. Correlation Analysis: Calculate the correlation between features and the target variable (house prices). Select features with high correlation.

2. Feature Importance: If you're using tree-based models (e.g., Random Forest or XGBoost), you can extract feature importance scores and select the most important features.

3. Recursive Feature Elimination: Use techniques like Recursive Feature Elimination (RFE) to iteratively remove less important features.

Model Training:

Choose the machine learning algorithms you want to use for your house price prediction. Common choices include Linear Regression, Random Forest, XGBoost, and Support Vector Machines. Here's a basic outline for model training:

1. Split the Data: Divide your dataset into training and testing sets to evaluate the model's performance.

2. Feature Scaling: Standardize or normalize your features, especially if you're using algorithms sensitive to feature scales.

3. Train the Models: Train different models using the training data. Tune hyperparameters to optimize model performance.

Evaluation:

Evaluating your model's performance is crucial to ensure it makes accurate predictions. Common evaluation metrics for regression tasks (like house price prediction) include:

1. Mean Absolute Error (MAE): Measures the average absolute difference between the predicted prices and the actual prices.

2. Mean Squared Error (MSE):Measures the average of the squared differences between predicted and actual prices.

3. Root Mean Squared Error (RMSE): The square root of MSE, which gives you an error metric in the same unit as the target variable.

4. R-squared (R2):Indicates the proportion of the variance in the dependent variable (house prices) that's predictable from the independent variables (features).

5. Cross-Validation: Perform cross-validation to assess the model's generalization performance.

6. Visualization: Use plots like scatter plots to visualize the actual vs. predicted prices

Program:

# Feature Selection

# Example: Using correlation analysis to select relevant features

correlation\_matrix = dataset.corr()

correlation\_with\_target = correlation\_matrix['HousePrice'] # Assuming 'HousePrice' is your target variable

relevant\_features = correlation\_with\_target[abs(correlation\_with\_target) > 0.2].index.tolist()

selected\_data = dataset[relevant\_features]

# Model Training

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

# Split the data into training and testing sets

X = selected\_data.drop('HousePrice', axis=1)

y = selected\_data['HousePrice']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Train a linear regression model

model = LinearRegression()

model.fit(X\_train, y\_train)

# Evaluation

from sklearn.metrics import mean\_squared\_error, r2\_score

# Make predictions

y\_pred = model.predict(X\_test)

# Calculate evaluation metrics

mse = mean\_squared\_error(y\_test, y\_pred)

r2 = r2\_score(y\_test, y\_pred)

# Print the results

print(f"Mean Squared Error: {mse}")

print(f"R-squared (R2): {r2}")

This code provides a basic example of feature selection, model training using Linear Regression, and evaluation using Mean Squared Error (MSE) and R-squared (R2) as metrics.

Conclusion:

Feature selection, model training, and evaluation form the core of effective house price prediction. These steps ensure that our models are robust and precise, empowering us to make informed decisions in the dynamic world of real estate.