**HOUSE PRICE PREDICTION**

**Phase-4 Document submission**

**TEAM MEMBER**

**NAME: RISWANA NASRIN PP**

**Reg.No:721921243087**

Project: House price prediction using ml

Phase 4: Development part 2

Topic:In this part we will continue building our project. Continue building the house price prediction model by Feature selection,Model training ,Evaluation

INTRODUCTION:

**House price prediction using machine learning is a cutting-edge application of artificial intelligence that leverages data analysis and predictive modeling to estimate the value of residential properties. By analyzing various factors such as location, square footage, number of bedrooms and bathrooms, and historical pricing trends, machine learning algorithms can provide accurate and data-driven predictions. In this field, data preprocessing, feature engineering, and model selection are crucial steps, making it an exciting and practical domain for both aspiring data scientists and real estate enthusiasts. This introduction will delve into the key concepts and methodologies used to forecast house prices with machine learning, offering valuable insights into this dynamic and ever-evolving field.**

Feature Selection:

**Feature selection involves choosing the most important variables (features) for predicting house prices**

**Identify relevant features, such as square footage, location, number of bedrooms, and more.**

**Use techniques like correlation analysis or feature importance scores to select the most influential features.**

**Eliminate irrelevant or redundant features to simplify the model**

SELECTING FEATURE

**L**ocation: The geographical location of the property can significantly impact its price. Features like neighborhood, city, or proximity to amenities can be important.

Square Footage: The size of the property, including the total square footage and individual room sizes, is a fundamental feature.

Number of Bedrooms and Bathrooms: The number of bedrooms and bathrooms is often a key factor influencing a house's value.

Lot Size: The size of the land on which the property is situated can be important, especially for houses with outdoor space.

Year Built: The age of the property can affect its value. Newer properties might be priced higher.

Amenities and Upgrades: Features like a pool, fireplace, modern kitchen appliances, or recent renovations can impact the price.

Local Schools: Proximity to good schools can be a crucial feature for families and can affect property values.

Crime Rates: Lower crime rates in the area can be a selling point and influence house prices.

LMarket Trends: Historical pricing trends in the local real estate market can also be used as a feature.

Economic Factors: Local economic conditions, employment opportunities, and interest rates can influence property values.

Accessibility: Proximity to public transportation, highways, and amenities like shopping centers can be important.

Property Type: Whether it's a single-family home, condominium, townhouse, or other property types can be a key feature.

Architectural Style: The architectural style and design of the house can influence its value.View: Properties with attractive views, such as waterfront or scenic vistas, can command higher prices.Energy Efficiency: Features like energy-efficient appliances, insulation, and solar panels can impact value.

Model Training:

**Model training is where you teach your machine learning algorithm to predict house prices.Select a suitable regression algorithm (e.g., linear regression, decision trees, random forests).Split your dataset into training and testing sets to assess the model's performance.Fit the model using the training data and fine-tune its parameters to improve accuracy.**

**TRAINING MY DATA**

Linear Regression: Linear regression is a straightforward and interpretable model. It assumes a linear relationship between the input features and the target variable (house price). It's a good starting point for simple cases.

Decision Trees: Decision trees can capture non-linear relationships and interactions between features. They are easy to understand and visualize. You can use ensemble methods like Random Forests to improve accuracy.

Gradient Boosting: Algorithms like Gradient Boosting (e.g., XGBoost, LightGBM) are powerful for predictive modeling. They can handle complex relationships and often provide high accuracy. They are also robust against overfitting.

Support Vector Regression (SVR): SVR is useful when there is a non-linear relationship between features and the target variable. It works well in high-dimensional spaces and is effective in capturing complex patterns.

Neural Networks: Deep learning models, such as feedforward neural networks, can be used for house price prediction, especially when you have a large dataset. They can automatically learn complex feature interactions.

K-Nearest Neighbors (KNN): KNN is a simple algorithm that considers the "k" closest data points to make predictions. It can work well when the prices of nearby properties are similar.

Ridge and Lasso Regression: These are variants of linear regression that can help with feature selection and regularization, reducing the risk of overfitting

ElasticNet Regression: ElasticNet combines L1 (Lasso) and L2 (Ridge) regularization. It's useful when you have many features, some of which may be irrelevant.

**Evaluation for house price prediction**

**Mean Absolute Error (MAE): MAE measures the average absolute difference between the actual house prices and the predicted prices. It gives you a sense of how far off, on average, your predictions are from the true values. A lower MAE indicates a more accurate model.Mean Squared Error (MSE): MSE measures the average squared difference between actual and predicted prices. It penalizes larger errors more than MAE. It provides a sense of how well your model handles outliers.Root Mean Squared Error**

(RMSE): RMSE is the square root of MSE. It provides a measure of the error in the same units as the target variable (house prices). RMSE is a common choice when you want to understand the magnitude of prediction errors

.R-squared (R²): R-squared quantifies the proportion of the variance in the target variable that is explained by your model. A higher R² indicates a better fit, with 1 being a perfect fit.

Adjusted R-squared: Adjusted R² considers the number of predictors in your model and adjusts R² to account for overfitting. It helps you determine whether adding more features is beneficial

Cross-Validation: Cross-validation techniques, such as k-fold cross-validation, help estimate how well your model will perform on unseen data. It assesses the model's robustness and generalization to new properties.

Residual Plots: Visualizing the residuals (the differences between actual and predicted prices) can reveal patterns and anomalies in your predictions. It's a valuable tool for identifying potential model shortcomings.

Comparing Models: To evaluate your model's performance, compare it to other models, industry benchmarks, or historical market data. You can use metrics like AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion) to compare different models

.Feature Importance Analysis: If you've used models like decision trees or random forests, you can analyze feature importance scores to understand which features have the most significant impact on predictions.Domain Expert

**Feedback**

Get feedback from real estate experts or professionals to assess if your model's predictions align with their expectations and market knowledge.