GRT INSTITUTE OF

ENGINEERING AND

TECHNOLOGY, TIRUTTANI - 631209

Approved by AICTE, New Delhi Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

PROJECT TITLE

HOUSE PRICE PREDICTION USING MACHINE LEARNING

COLLEGE CODE:1103

JAGADEESH.S

3rd yr, 5th sem

Reg no.:110321104016

Jagadeeshpu4lyf@gmail.com

PROJECT NAME

HOUSE PRICE PREDICTION USING MACHINE LEARNING

SUMMITED BY : Jagadeesh S

MAIL ID : jagadeeshpu4lyf@gmail.com

AIM:

- People looking to buy a new home tend to be more conservative with their budgets and market strategies.
- This project aims to analyses various parameters like average income, average area etc. and predict the house price accordingly.
- This application will help customers to invest in an estate without approaching an agent
- To provide a better and fast way of performing operations.
- To provide proper house price to the customers.
- To eliminate need of real estate agent to gain information regarding house prices.
- To provide best price to user without getting cheated.
- To enable user to search home as per the budget
- The aim is to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. By analyzing previous market trends and price range , and also upcoming developments future prices will be predicted.
- House prices increase every year, so there is a need for a system to predict house prices in the future.
- House price prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house.
- We use linear regression algorithm in machine learning for predicting the house price trends

ABSTRACT:

Data mining is now commonly applied in the real estate market. Data mining's ability to extract relevant knowledge from raw data makes it very useful to predict house prices, key housing attributes, and many more. Research has stated that the fluctuations in house prices are often a concern for house

owners and the real estate market. A survey of literature is carried out to analyse the relevant attributes and the most efficient models to forecast the house prices. The findings of this analysis verified the use of the Artificial Neural Network, Support Vector Regression and XG Boost as the most efficient models compared to others. Moreover, our findings also suggest that locational attributes and structural attributes are prominent factors in predicting house prices. This study will be of tremendous benefit, especially to housing developers and researchers, to ascertain the most significant attributes to determine house prices and to acknowledge the best machine learning model to be used to conduct a study in this field

INTRODUCTION:

House is one of human life's most essential needs, along with other fundamental needs such as food, water, and much more. Demand for houses grew rapidly over the years as people's living standards improved. While there are people who make their house as an investment and property, yet most people around the world are buying a house as their shelter or as their livelihood.

The real estate industry is one of the most significant sectors of the global economy, and for both homebuyers and sellers, accurately predicting house prices is of paramount importance. Traditionally, real estate professionals have relied on their expertise and market knowledge to estimate property values. However, the advent of machine learning has revolutionized the way we approach this task.

PROBLEMS IN HOUSE PRICE PREDICTION:

Predicting house prices using machine learning can be a challenging task due to various factors and potential problems. Here are some common issues and challenges associated with house price prediction using machine learning

Data Quality:

Incomplete data: Missing values in the dataset can lead to biased predictions.

Noisy data: Outliers and errors in the data can negatively impact model accuracy.

Feature Selection and Engineering:

Choosing the right features (variables) and transforming them appropriately is crucial. Including irrelevant features or excluding important ones can affect model performance.

Creating meaningful features from raw data, such as converting categorical variables into numerical representations (e.g., one-hot encoding), can be complex.

Over fitting:

Over fitting occurs when a model learns to perform well on the training data but fails to generalize to unseen data. This can happen if the model is too complex relative to the amount of training data available.

Under fitting:

Under fitting occurs when a model is too simple to capture the underlying patterns in the data. This can happen if the model lacks complexity or if it is not trained for a sufficient number of epochs.

Data Imbalance:

Imbalanced data, where one class (e.g., expensive houses) is underrepresented, can lead to biased predictions.

Scalability:

Scalability can be an issue when dealing with a large dataset. Training a machine learning model on a massive dataset can be computationally expensive and time-consuming.

Non-Linearity:

House price prediction often involves non-linear relationships between features and the target variable. Simple linear regression models may not capture these relationships effectively.

Model Selection:

Choosing the right machine learning algorithm or model architecture can be challenging. Different algorithms have different strengths and weaknesses, and selecting the wrong one can lead to suboptimal results.

Hyper parameter Tuning:

Setting the hyper parameters (e.g., learning rate, regularization strength) of a model is essential for achieving good performance. Tuning these hyper parameters can be time-consuming and require domain expertise.

Data Leakage:

Data leakage occurs when information from the test set inadvertently leaks into the training set, leading to overly optimistic performance estimates. Proper data splitting and preprocessing are crucial to prevent this.

Market Dynamics:

House prices can be influenced by various factors such as economic conditions, location-specific trends, and seasonality. Capturing these external factors in the model can be challenging.

Interpretability:

Some machine learning models, such as deep neural networks, can be difficult to interpret. This can be a problem when stakeholders require explanations for the predictions.

Ethical and Fairness Concerns:

Biases in the data or model can lead to unfair predictions, such as discriminating against certain groups. Ensuring fairness in housing predictions is important.

WAYS TO FIX THOSE PROBLEMS:

Fixing the problems associated with house price prediction using machine learning involves a combination of data pre-processing, model selection, and tuning. Here are steps you can take to address these problems.

Data Quality:

Address missing data: You can use techniques like imputation (e.g., filling missing values with means or medians) or, if applicable, collect more data to reduce missing values.

Handle outliers:

Identify and remove or transform outliers in the data to prevent them from skewing the model's predictions.

Feature Selection and Engineering:

Conduct feature selection: Use techniques like feature importance scores, recursive feature elimination, or domain knowledge to select the most relevant features.

Engineer new features:

Create meaningful features that capture the underlying patterns in the data. For example, you can calculate the price per square foot or include time-based features for seasonality.

Over fitting and Under fitting:

Regularization: Apply regularization techniques like L1 (Lasso) or L2 (Ridge) regularization to prevent over fitting.

Cross-validation:

Use techniques like k-fold cross-validation to assess your model's generalization performance and fine-tune hyper parameters accordingly.

Data Imbalance:

If dealing with imbalanced data, use techniques such as oversampling the minority class or under sampling the majority class to balance the dataset.

Scalability:

Consider distributed computing frameworks like Apache Spark or cloud-based solutions for handling large datasets efficiently.

Non-Linearity:

Use non-linear models like decision trees, random forests, gradient boosting, or neural networks to capture complex relationships in the data.

Model Selection:

Experiment with different machine learning algorithms to find the one that performs best on your specific dataset. Consider ensemble methods for improved accuracy.

Hyper parameter Tuning:

Utilize automated hyper parameter optimization techniques such as grid search, random search, or Bayesian optimization to find the best hyper parameter settings for your model.

Data Leakage:

Carefully split your data into training, validation, and test sets to prevent data leakage. Ensure that any pre-processing steps are applied separately to each dataset.

Market Dynamics:

Incorporate external factors like economic indicators, location-specific trends, and seasonality into your model if they are relevant to house price prediction.

Interpretability:

Choose models that offer better interpretability, such as linear regression or decision trees, when transparency is essential.

Ethical and Fairness Concerns:

Evaluate your model for bias and fairness and consider techniques like re-sampling, re-weighting, or adversarial de biasing to mitigate bias in predictions.

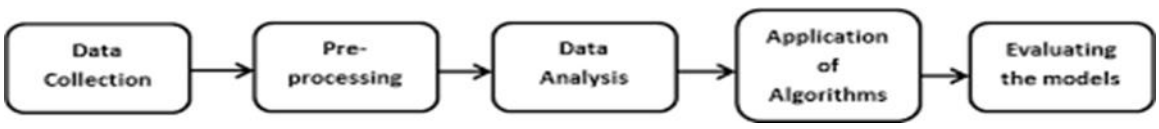
Continuous Monitoring and Improvement:

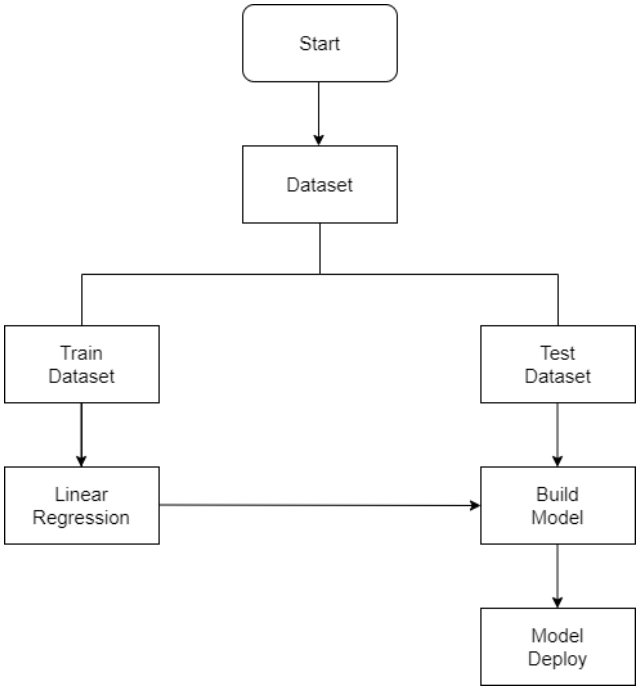
House price prediction models should be updated regularly to account for changing market conditions. Monitor model performance over time and retrain it as needed.

Domain Expertise:

Collaborate with domain experts or real estate professionals who can provide insights into the factors affecting house prices in your target market.

DESIGN:





CONCLUSION:

developing a house price prediction model using machine learning is a complex but rewarding endeavor. By systematically addressing the common challenges and problems associated with this task, we can create a robust and accurate predictive model. Here's a summary of the key takeaways.

- **Data Quality Matters**
- **Feature Engineering**
- **Over fitting and Under fitting**
- **Imbalanced Data**
- **Scalability**
- **Model Selection**
- **Hyper parameter Tuning**
- **Data Leakage**
- **Market Dynamics**
- **Interpretability**
- **Ethical Considerations**
- **Continuous Monitoring and Improvement**

successfully predicting house prices using machine learning involves a combination of data pre-processing, thoughtful model selection, rigorous evaluation, and on-going maintenance. By addressing these challenges and taking a systematic approach, you can build a reliable and valuable tool for the real estate market, benefiting both buyers and sellers. Moreover, a well-documented and transparent process ensures accountability and trust in your predictive model.