**HOUSE PRICE PREDICTOR**

**INTRODUCTION:**

In the realm of data science and machine learning, predicting house prices is a classic and highly valuable task. Our project embarks on this journey, leveraging a rich array of tools, techniques, and technologies to create a predictive model for house prices. Through this endeavor, we aim to demonstrate a comprehensive understanding of data science concepts, from data preprocessing to model deployment.

**PROBLEM STATEMENT:**

The housing market is an important and complex sector that impacts people's lives in many ways. For many individuals and families, buying a house is one of the biggest investments they will make in their lifetime. Therefore, it is essential to accurately predict the prices of houses so that buyers and sellers can make informed decisions. This project aims to use machine learning techniques to predict house prices based on various features such as location, square footage, number of bedrooms and bathrooms, and other relevant factors.

**PROBLEM DEFINITION:**

The problem is to predict house prices using machine learning techniques. The objective is to develop a model that accurately predicts the prices of houses based on a set of features such as location, square footage, number of bedrooms and bathrooms, and other relevant factors. This project involves data preprocessing, feature engineering, model selection, training, and evaluation.

**DESIGN THINKING:**

We will first build a model using sklearn and linear regression using banglore home prices dataset from kaggle.com. Second step would be to write a python flask server that uses the saved model to serve http requests. Third component is the website built in html, css and javascript that allows user to enter home square ft area, bedrooms etc and it will call python flask server to retrieve the predicted price. During model building we will cover almost all data science concepts such as data load and cleaning, outlier detection and removal, feature engineering, dimensionality reduction, gridsearchcv for hyperparameter tunning, k fold cross validation etc.

Technology and tools wise this project covers

1.**Data source:** kaggle.com

Kaggle is an online platform that is popular among data scientists, machine learning practitioners, and data enthusiasts.

2**. Data preprocessing**: Numpy and Panda

NumPy is a fundamental library for numerical and scientific computing in Python. It is commonly used for data preprocessing, mathematical operations on data, and creating data structures to work with in machine learning and data analysis.

3. **Feature selection**: Regression Feature Elimination from sklearn library

RFE is a technique that recursively removes less important features from the dataset and fits the model with the remaining features. It assigns weights to each feature and eliminates the ones with the lowest weights. This process is repeated until the desired number of features is reached.

4. **Model selection**: Linear regression

Linear regression is a fundamental statistical and machine learning technique used for modeling the relationship between a dependent variable (target) and one or more independent variables (features) by fitting a linear equation to the observed data.

5**. Model training** : Sklearn

It offers a comprehensive set of tools and functionalities for building, training, and evaluating machine learning models.

6. **Evaluation**: K- fold cross validation

K-fold cross-validation is a technique used to assess the performance of a machine learning model by dividing the dataset into "K" subsets (folds). The model is trained and evaluated "K" times, each time using a different fold as the validation set and the remaining folds as the training set. This allows you to obtain multiple performance metrics and assess the model's generalization ability.

**CONCLUSION**:

The project is a comprehensive example of a data science project that involves various stages, from data preprocessing to model building and deployment. It demonstrates the application of data science concepts and tools to solve a real-world problem of predicting house prices, with a strong emphasis on both model development and web-based user interaction.