TITLE: House Price Prediction App

Overview:

The value of real estate continues to be a focal point for homeowners, investors, and the housing market at large. Accurate predictions of house prices play a pivotal role in informed decision-making. Accordingly, this project harnesses machine learning methodologies to develop a predictive model for house prices, incorporating user-friendly web interfaces via Streamlit. Central to this endeavour is the Ames Housing Dataset, a widely recognised benchmark in the domain of housing price prediction.

Aim:

The principal aim of this project is to provide an effective tool for predicting the selling price of individual houses based on specified features. By leveraging machine learning techniques, this project seeks to enhance the accuracy of price estimations, thereby benefitting potential buyers and sellers alike.

Features:

The model focuses on key features that significantly influence house prices, such as:

1. Square Footage: Total area of the property.

2. Number of Bedrooms:A count of the bedrooms in the house.

3. Number of Bathrooms:Total number of bathrooms, including full and half.

4. Location:Geographical area or neighbourhood, which affects desirability.

5. Year Built: The age of the property, often impacting value.

6. Garage Size: Dimensions of available parking space and its capability to house vehicles.

These features are pivotal in defining the algorithm's input and enhancing model accuracy.

Purpose:

The primary purpose of this project is twofold: first, to develop a robust predictive model that can yield credible price predictions, and second, to create an intuitive web application using Streamlit that empowers users to interactively input data and receive instantaneous price estimates. This

alignment of advanced machine learning with modern web technologies ensures an accessible tool for a diverse audience.

Procedure:

The procedural approach adopted in this project encompasses the following stages:

1**. Data Collection and Preprocessing**:Extracting the relevant features from the Ames Housing Dataset and applying thorough data cleaning to handle missing values and categorical variables.

2. **Model Selection and Training**: Employing a Random Forest Regressor due to its ensemble nature and superior performance on regression tasks. The training phase entails splitting the data into training and testing sets to allow for model validation.

3. **Model Evaluation**:Utilising metrics such as Mean Absolute Error (MAE) and R-squared to assess and optimise model performance.

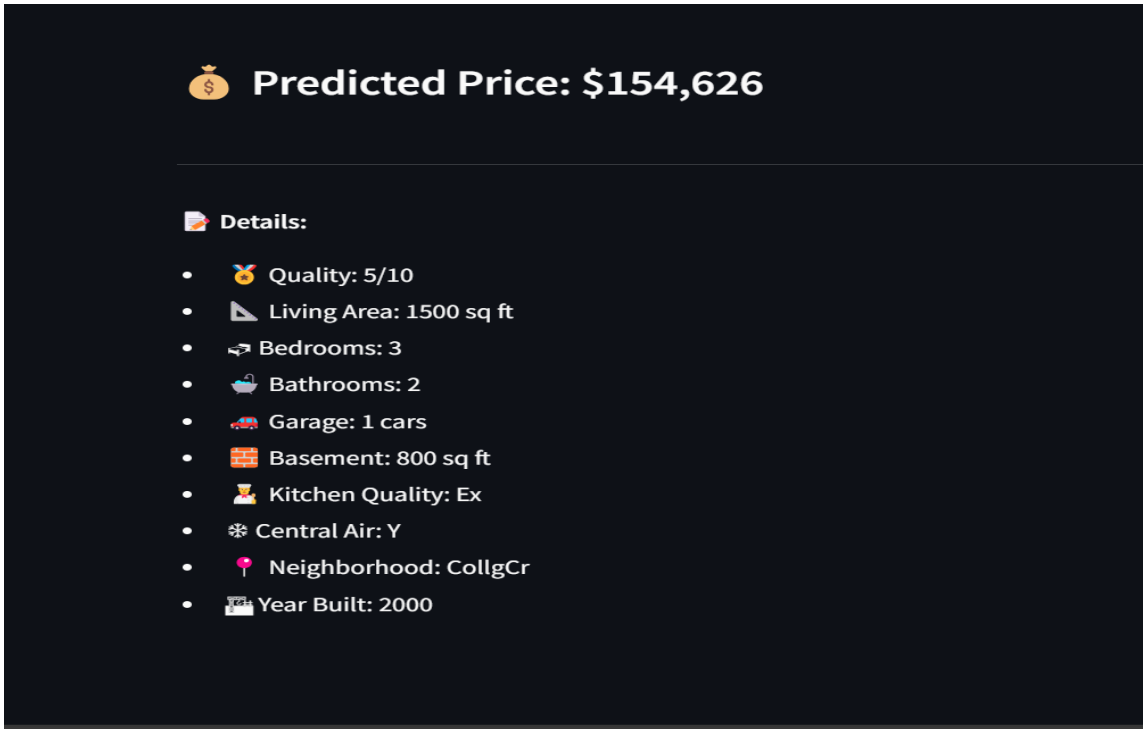
4. **Web Application Development**: Implementing Streamlit to create an interactive web interface that enables real-time user input, processing that input through the model, and providing immediate feedback on predicted prices.

Code:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.ensemble import RandomForestRegressor
import joblib
df = pd.read_csv("AmesHousing.csv")
features = [
    "Overall Qual",
    "Gr Liv Area",
    "Bedroom AbvGr",
    "Full Bath",
    "Garage Cars",
    "Total Bsmt SF",
    "Kitchen Qual",
    "Central Air",
    "Neighborhood",
    "Year Built"
]
target = "SalePrice"
df = df[features + [target]].dropna()
X = df.drop(target, axis=1)
y = df[target]
categorical = ["Neighborhood", "House Style", "Exterior 1st", "Kitchen Qual"]
numeric = [col for col in X.columns if col not in categorical]
preprocessor = ColumnTransformer([
    ("cat", OneHotEncoder(drop="first", handle_unknown='ignore'), categorical)
], remainder="passthrough")
```

Output:


House Price Predictor
Enter House Details
Overall Quality (1-10)
5
Above Ground Living Area (sq ft)
1500
Bedrooms Above Ground
3
Full Bathrooms
2
Garage Capacity (Cars)
1


Predicted Price: $154,626
Details:
Quality: 5/10
Living Area: 1500 sq ft
Bedrooms: 3
Bathrooms: 2
Garage: 1 cars
Basement: 800 sq ft
Kitchen Quality: Ex
Central Air: Y
Neighborhood: CollgCr
Year Built: 2000

Conclusion:

In summary, this project successfully merges machine learning with accessible web-based applications to address the challenging task of house price prediction. By employing a Random Forest Regressor trained on the Ames Housing Dataset, it affords potential buyers and sellers a reliable tool for understanding market valuations. With further enhancement in feature selection and exploratory data analysis, the model holds the potential for even greater accuracy and user engagement. As the housing market continues to evolve, the application of such predictive modelling will undoubtedly remain invaluable in guiding decision-making processes within this domain.