# **INTRODUCTION** :



* The real estate market is one of the most dynamic and lucrative sectors, with house prices constantly fluctuating based on various factors such as location, size, amenities, and economic conditions. Accurately predicting house prices is crucial for both buyers and sellers, as it can help make informed decisions regarding buying, selling, or investing in properties.
* Traditional linear regression models are often employed for house price prediction. However, they may not capture complex relationships between predictors and the target variable, leading to suboptimal predictions. In this project, we will explore advanced regression techniques to enhance the accuracy and robustness of house price prediction models.
* Briefly introduce the real estate market and the importance of accurate house price prediction.

Highlight the limitations of traditional linear regression models in capturing complex relationships.

* Emphasize the need for advanced regression techniques like gradient boosting and XGBoost to enhance prediction accuracy

Content for Project Phase 2 :

Consider exploring advanced regression techniques like Gradient Boosting or XGBoost for

improved Prediction accuracy.

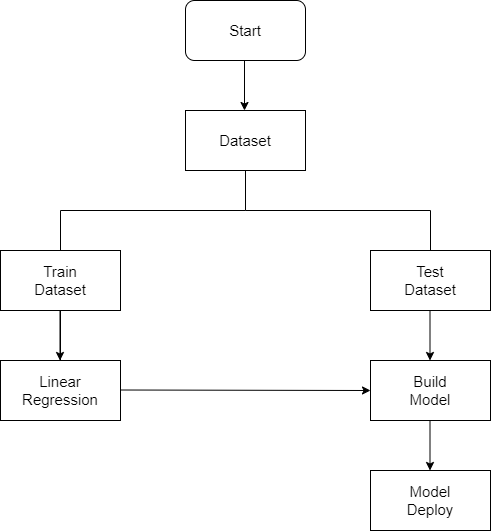
Data Source

A good data source for house price prediction using machine learning should be

Accurate, Complete, Covering the geographic area of interest, Accessible.

Dataset Link: (<https://www.kaggle.com/datasets/vedavyasv/usa-housing>)

Flow chart of dataset:



* SOFTWARETOOL
  + Jupyter
  + keras
  + Visual studio
  + R Square
  + Adjusted R Square
  + MSE
  + RMSE
  + MAE
  + GOOGLE COLLA

**PROPOSED SYSTEM PHASES**

**Phase 1: Collection of data**

* Data processing techniques and processes are numerous. We collecteddata for USA/Mumbai

real estate properties from various real estate websites.

* The data would be having attributes such as Location, carpet area, built-up area, age of the property, zip code, price, no of bedrooms etc.
* We must collect the quantitative data which is structured and categorized. Data

collection is needed before any kind of machine learning research is carried out. Dataset

validity is a must otherwise there is no point in analyzing the data.

**Phase 2: Data preprocessing**

* Data preprocessing is the process of cleaning our data set. There might be missing values or

outliers in the dataset. These can be handled by data cleaning. If there are many missing

values in a variable we will drop those values or substitute it with the average value.

**Phase 3: Training the model**

* Since the data is broken down into two modules: a Training set and Test set, we must initially

train the model. The training set includes the target variable.

* The decision tree regression algorithm is applied to the training data set. The Decision tree builds a regression model the form of a tree structure.

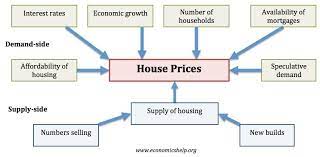
**Phase 4: Testing and Integrating with UI**

* The trained model is applied to test dataset and house prices are predicted. The trained model is then integrated with the front end using Flask in python

**FACTORS THAT AFFECT HOUSE PRICING**

In order to predict house prices, first we have to understand the factors that affect house

pricing.



* + **Economic growth**. Demand for housing is dependent upon income. With higher economic growth and rising incomes, people will be able to spend more on houses; this will increase demand and push up prices. In fact, demand for housing is often noted to be income elastic (luxury good); rising incomes leading to a bigger % of income being spent on houses. Similarly, in a recession, falling incomes will mean people can’t afford to buy and those who lose their job may fall behind on their mortgage payments and end up with their home repossessed.
  + **Unemployment**. Related to economic growth is unemployment. When unemployment is rising,fewer people will be able to afford a house. But, even the fear of unemployment may discourage people from entering the property market.
  + **Interest rates**. Interest rates affect the cost of monthly mortgage payments. A period of high- interest rates will increase cost of mortgage payments and will relatively more attractive compared to buying. Interest rates have a bigger effect if homeowners have large variable mortgages. For example, in 1990-92, the sharp in interest rates caused a very steep fall in UK house prices because many homeowners couldn’t afford the rise in interest rates.
  + **Consumer confidence**. Confidence is important for determining whether people

want to take therisk of taking out a mortgage. In particular expectations towards

the housing market is important; if people fear house prices could fall, people will

defer buying.

* + **Mortgage availability**. In the boom years of 1996-2006, many banks were very

keen to lend mortgages. They allowed people to borrow large income multiples

(e.g. five times income). Also, banks required very low deposits (e.g. 100%

mortgages). This ease of getting a mortgage meant that demand for housing

increased as more people were now able to buy. However, since the credit crunch

of 2007, banks and building societies struggled to raise funds for lending on the

money markets. Therefore, they have tightened their lending criteria requiring a

bigger deposit to buy a house. This has reduced the availability of mortgages and

demand fell.

* + **Supply**. A shortage of supply pushes up prices. Excess supply will cause prices to

fall. For example, inthe Irish property boom of 1996-2006, an estimated 700,000

new houses were built. When the property market collapsed, the market was left

with a fundamental oversupply. Vacancy rates reached 15%, and with supply

greater than demand, prices fell.

* **Geographical factors.** Many housing markets are highly geographical. For example,

national house prices may be falling, but some areas (e.g. London, Oxford) may

still see rising prices. Desirable areas can buck market trends as demand is high,

and supply limited. For example, houses near goodschools or a good rail link may

have a significant premium to other areas. This graph shows that first time buyers

in London face much more expensive house prices – over 9.0 times earnings

compared to the north, where house prices are only 3.3 times earnings.

SAMPLE CODE

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

%matplotlib inline

HouseDF = pd.read\_csv('USA\_Housing.csv')

HouseDF.head()

HouseDF=HouseDF.reset\_index()

HouseDF.head()

HouseDF.info()

HouseDF.describe()

HouseDF.columns

sns.pairplot(HouseDF)

sns.distplot(HouseDF['Price’])

sns.heatmap(HouseDF.corr(), annot=True)

X = HouseDF[['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms','Avg. Area

Number of Bedrooms', 'Area Population']]

y = HouseDF['Price’]

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

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y,test\_size=0.4,random\_state=101)

from sklearn.linear\_model import minmaxscaler

lm = minmaxscaler(feature\_range=(0,1))

lm.fit\_transform(X\_train,y\_train)

print(lm.intercept\_)

coeff\_df = pd.DataFrame(lm.coef\_,X.columns,columns=['Coefficient’])

coeff\_df

from keras.layersimport Dense,Dropout,LSTM

from keras.models import Sequential

model = Sequential()

model.add(LSTM(units = 50,activation = 'relu',return\_sequences = True,input\_shape =

(x\_train.shape[1], 1)))

model.add(Dropout(0.2))

model.add(LSTM(units = 60,activation = 'relu',return\_sequences = True))

model.add(Dropout(0.3))

model.add(LSTM(units = 80,activation = 'relu',return\_sequences = True))

model.add(Dropout(0.4))

model.add(LSTM(units = 120,activation = 'relu'))

model.add(Dropout(0.5))

model.add(Dense(units = 1))

model.compile(optimizer='adam', loss = 'mean\_squared\_error’)

model.fit(x\_train, y\_train,epochs=50)

print(lm.intercept\_)

coeff\_df = pd.DataFrame(lm.coef\_,X.columns,columns=['Coefficient’])

coeff\_df

predictions = lm.predict(X\_test)

scale\_factor = 1/0.02099517

y\_predicted = y\_predicted \* scale\_factory

y\_test = y\_test \* scale\_factor

plt.scatter(y\_test,predictions)

sns.distplot((y\_test-predictions),bins=50);

plt.figure(figsize=(12,6))

plt.plot(y\_test,'b',label = 'Original Price')

plt.plot(y\_predicted,'r',label = 'Predicted Price')

plt.xlabel('Time')

plt.ylabel('Price')

plt.legend()

plt.show()

from sklearn import metrics

print('MAE:', metrics.mean\_absolute\_error(y\_test, predictions))

print('MSE:', metrics.mean\_squared\_error(y\_test, predictions))

print('RMSE:', np.sqrt(metrics.mean\_squared\_error(y\_test, predictions)))

**ADVANTAGE OF LSTM OVER OTHER MODELS**

The LSTM model can be tuned for various parameters such as changing the number of LSTM layers, adding dropout value or increasing the number of epochs.

Long Short Term Memory (LSTM)

LSTMs are widely used for sequence prediction problems and have proven to be

extremely effective. The reason they work so well is because LSTM is able to store past

information that is important, and forget the information that is not. LSTM has three

Gates:

The input gate: The input gate adds information to the cell state

The forget gate: It removes the information that is no longer required by the model. The

output gate: Output Gate at LSTM selects the information to be shown as output

CONCLUSION

Thus the machine learning model to predict the house price based on given dataset is executed successfully using xg regressor (a upgraded/slighted boosted form of regular linear regression, this gives lesser error). This model further helps people understand whether this place is more suited for them based on heatmap correlation. It also helps people looking to sell a house at best time for greater profit. Any house price in any location can be predicted with minimum error by giving appropriate dataset.

**Project Conclusion:**

In the Phase 2 conclusion, we will summarize the key findings and insights from the advanced regression techniques. We will reiterate the impact of these techniques on improving the accuracy and robustness of house price predictions.

* + Future Work: We will discuss potential avenues for future work, such as incorporating additional data sources (e.g., real-time economic indicators), exploring deep learning models for prediction, or expanding the project into a web application with more features and interactivity.