**PROJECT: PREDICTING HOUSE PRICE USING MACHINE LEARNING  
PHASE 3: DEVELOPEMENT PROCESS 1**

**OUTLINE: -**

* Project Summary
* Technology used
* Tools used
* How it works?
* What it does?

**INTRODUCTION:-**

**Problems faced during buying a house:**

1. Buying a house is a stressful thing.
2. Buyers are generally not aware of factors that influence the house prices.
3. Many problems are faced during buying a house.
4. Hence real estate agents are trusted with the communication between buyers and sellers as well as laying down a legal contract for the transfer. This just creates a middle man and increases the cost of houses.

**They believe that it depends upon:**

1. The square foot area
2. Neighbourhood
3. The number of bedrooms

**But it depends upon many factors also…. such as:**

1. No.of storeys
2. Area outside the house
3. Rooms on one floor

**PROJECT SUMMARY:**

* Our project is a machine learning app, based on certain specifications of your future home it will try to guess the most accurate price.
* Information such as state, city, area, stores.

**TECHNOLOGY USED:**

1. **Machine Learning:**

* “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if it performance at tasks in T, as measured by P, improves with experience E.”
* Example: predicting whether the given object is pen or pencil?

E=the experience or predicting many pens and pencil

T=the task of predict pen or pencil

P=the probability that of whether it is a pen or pencil.

In general, any machine learning problem can be assigned to one of the two broad classifications:

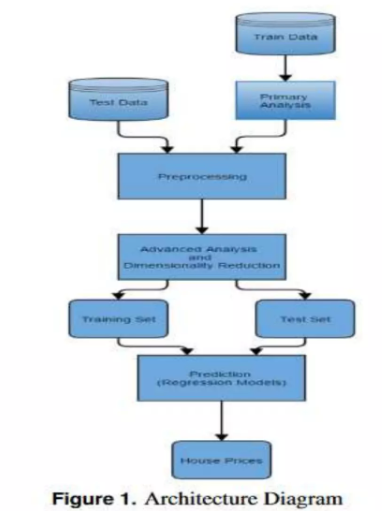
* Supervised learning .
* Unsupervised learning.

**TOOLS USED:**

* PYTHON
* TENSOR FLOW
* Android Studio
* Anaconda

**HOW IT WORKS?**

* Collecting the data: First step was to collect data we collected data from different sources and merged them together to form our training data set.
* Then we trained the model using machine learning algorithm which in this case is multiple linear regression.
* Based on the generated graphs we predict the cost of the house.



**FUTURE WORK:**

* Our model had a low rmse score, but there is still room for improvement. In a real world scenario, we can use such a model to predict house prices. This model should check for new data, once In a month, and incorporate them to expand the dataset and produce better results.
* We can try out other dimensionality reduction techniques like Univariate feature selection and Recursive feature elimination in the initial stage.
* We can try out other advanced regression techniques, like random forest and Bayesian ridge algorithm, for prediction. Since the data is highly correlated, we should also try Elastic Net regression technique.

**PROPOSED SYSTEM: -**

* System includes set of codes that processes on the available dataset to effectively predicts the value of outcome depending upon user input using the concept of Linear regression.
* Efficient and proper use of the system can eradicate the cases where the customers get cheated by the real estate agents in terms of house prices.
* Proper usage of model is beneficial to both the customer as well as agents guiding customers.

**ADVANTAGES: -**

* Good interpretability
* Its very simple to understand
* Space complexity is very low it just needs to save the weights at the end of training. It’s a high latency algorithm.

**Dataset description:**

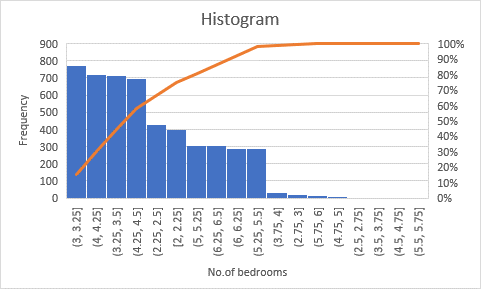
Dataset: [("C:\Users\Thahira\Desktop\MyProject\HousePricePrediction.xlsx")](file:///C:\Users\Thahira\Desktop\MyProject\USA_housing.csv)

Following are the parameters in dataset for house price prediction of ‘USA’

* Avg. Area Income
* House Age
* Number of Rooms
* Avg. Area Number of Bedrooms
* Area Population
* Price
* Address
* 24\*7 security

**VISUALISATION(SCATTER PLOT):-**

**DATA VISUALISATION(HISTOGRAM):-**



**APPLYING MACHINE LEARNING ALGORITHM:-**

* We set independent and target variables as x and y respectively.
* Split the dataset into training and testing in 70:30 ratio.
* Fitting the train set to multiple linear regression and getting the score of training model.
* Fitting the train set to decision tree and getting the score of the model.
* Fitting the train set to random forest and getting the score of the model.
* Calculate the model score to understand how our model performed along with the explained variance score.

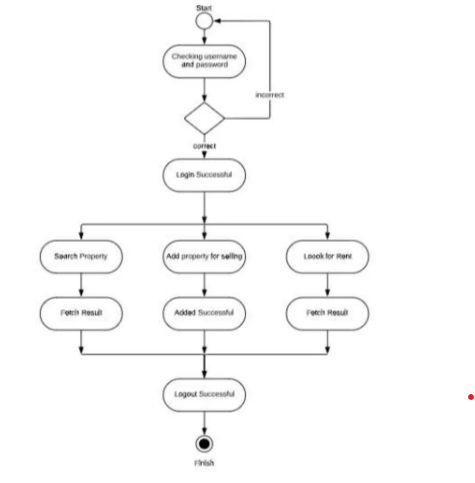
**UML DIAGRAMS:-**

* Use case diagrams are used to gather the requirements of a system including internal and external influences.

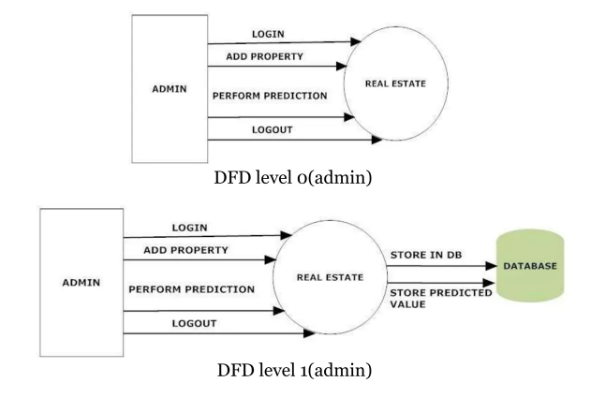
A diagram of a person using a case diagram

Description automatically generated

* Activity diagram is basically a flowchart to represent the flow from one activity to another activity



* A data-flow diagram(DFD) is a way of representing a flow of a data of a process or a system (usually an information system). The DFD also provides information about the outputs and inputs of each entity and the process itself.



* Sequence diagram emphasizes on sequence of messages.

A diagram of a web application

Description automatically generated

**METHOD OF ANALYSIS-RANDOM FOREST**

* + Variables
    - Using the same variable used in the linear regression.
* Trees and Nodes
* Checking various combinations of number of trees and maximum number of nodes to get the best result.
* Using number of trees=100 and maximum nodes=10 for best fitted model.
* Model Accuracy
* Checking model accuracy using error rate and MAPE.
* Decision
* Drop the variable if the model accuracy falls or remains same.

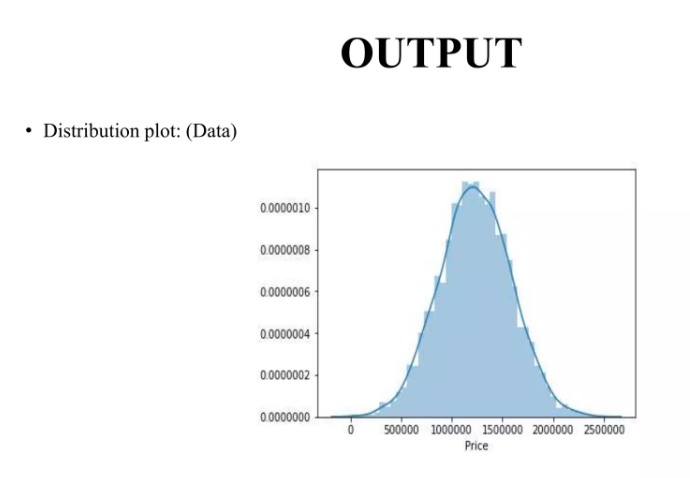
**FEASIBILITY STUDY:-**

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. The feasibility study of the proposed system is carried out. It is carried out to ensure that the proposed system is not a burden to the company.

* Economic feasibility
* Technical feasibility
* Social feasibility

A white background with black text

Description automatically generated



**A screenshot of a graph

Description automatically generated**

**CONCLUSION:**

Thus, we studied and applied the concept of Linear Regression in real time implementation so as to ease the life of human.

Determining the price of property without complete knowledge about the surrounding is quite riskier for both customer and the seller.

In order to overcome this problem we have tried to develop application which determines the price of the property based on various parameters of the surrounding.

• Data provide us with the complete data about the surrounding in the form of dataset.

• Dataset helps to get the insight of the surrounding and machine learning model helps to predict the price of the property based on the training provided by the dataset.

• We successfully implemented linear regression model to predict the price of the houses.