# INTRODUCING – "PROPERTY VALUE FORECASTING"

## Prof. Karande M.A*1, Ms. Tugavkar Trupti*2, Ms. Gade Amruta*3,
## Ms. Andhare Ketaki*4, Ms. Jikare Ayesha*5

*1Guide, Department Of Computer Science JSPM's Bhagwant Institute Of Technology, Barshi, India.

*2,3,4,5Department Of Computer Science  JSPM's Bhagwant Institute Of Technology, Barshi, India.

## ABSTRACT

The below document presents the implementation of price prediction project for the real estate markets and housing. Many algorithms are used here to effectively increase the accuracy percentage, various researchers have done this project and implemented the algorithms like hedonic regression, artificial neural networks, AdaBoost, J48 tree which is considered as the best models in the price prediction. These are considered as the base models and by the help of advanced data mining tools algorithms like a random forest, gradient boosted trees, multilayer perceptron and ensemble learning models are used and prediction accuracy is attained in a higher rate. The results and evaluation of these models using the machine learning and advanced data mining tools like Weka, Rapid Miner will have the more influence in the price prediction.

**Keywords**: Machine Learning, House Price Prediction, Real Estate, Python.

## I.     INTRODUCTION

Machine Learning (ML) is a vital aspect of present-day business and research. It progressively improves the performance of computer systems by using algorithms and neural network models. Machine Learning algorithms automatically build a mathematical model using sample data – also referred to as "training data" which form decisions without being specifically programmed to make those decisions.

ML techniques can be broadly categorized as:

1.  Supervised Learning- Here, the learning process is guided. The available dataset is used to train the built model or machine. Once trained, it is able to make predictions or take decisions when any new data is input to it.

2.  Unsupervised Learning- Here, the model studies via observations and notes the formations in the statistics. Once a dataset is provided to the model, it automatically learns samples and relationships in the data by creating clusters out of it. For instance, if images of bananas, mangoes, and apples are presented to the model, it creates clusters of the dataset based on some relationships and patterns, and segregates the images into those clusters. So, when new images are fed to the trained model, it can add it to one of the formed clusters.

3.  Reinforcement Learning- It refers to an agent's potential to interrelate with its surrounding environment and learn about the best possible outcome available. It goes around with the hit-and-trial theory wherein the agent's either rewarded or penalized with a point for each correct or wrong answer respectively. Then, basis the positive reward points obtained, the model directs itself. Once it gets upskilled, it becomes ready to anticipate the new data given to it.

## II.     LITERATURE REVIEW

A Litrature review on real estate price prediction would encompass various studies and methodologies aimed at forecasting property prices. Here's a structured outline for such a review:

**1.  Literature Review: Property Value Forecasting**

Introduction:

Define the importance of real estate price prediction in various contexts such as investment, urban planning, and housing policy. Highlight the significance of accurate price forecasting for stakeholders like buyers, sellers, investors, and policymakers.

## III.    METHODOLOGY AND IMPLEMENTATION

The below passages describe about the methodology used in the real estate house price predictions and the architecture flow diagram is given.
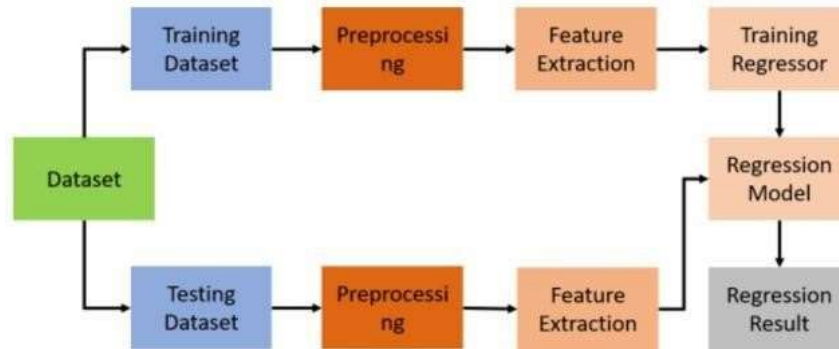


## IV.    MODELING AND ANALYSIS

**Property Value Forecasting:**

These features are given to the ML model and based on how these features affect the label it gives out a prediction.
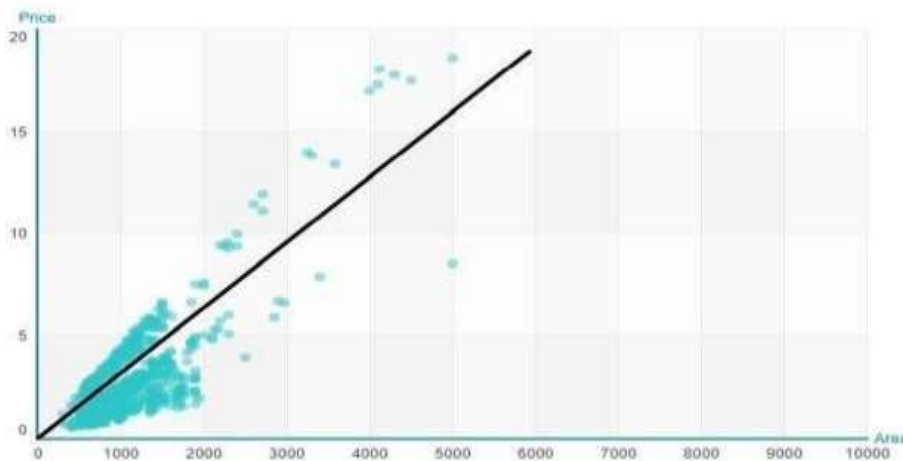


**Fig 1.** Linear Regression Scatter Plot.

Furthermore, after finalizing the dataset, the dataset will go through the process known as data cleaning where all the data which is not needed will be eliminated and the raw data will be turned into a .csv file. Moreover, the data will go through data preprocessing where missing data will be handled and if needed label encoding will be done. Moreover, this will go through data transformation where it will be converted into a NumPy array so that it can finally be sent for training the model. While training various machine learning algorithms will be used to train the model their error rate will be extracted and consequently an algorithm and model will be finalized which can yield accurate predictions.

**Employing Machine Learning for Property Value Forecasting:**

1. Data Refining: Data refinement means ensuring the data put into a data analytics platform is relevant, homogenized and categorized so the users can get meaningful results and pinpoint discrepancies. The data refinement process is a key part of establishing a data-driven company and maintaining good habits.

2. Regression: Regression analysis is a powerful statistical method that allows you to examine the relationship between two or more variables of interest. While there are many types of regression analysis, at their core they all examine the influence of one or more independent variables on a dependent variable.

3. Classification: Regression analysis is a powerful statistical method that allows you to examine the relationship between two or more variables of interest. While there are many types of regression analysis, at their core they all examine the influence of one or more independent variables on a dependent variable.

4. Clustering: Regression analysis is a powerful statistical method that allows you to examine the relationship between two or more variables of interest. While there are many types of regression analysis, at their core they all examine the influence of one or more independent variables on a dependent variable.

**Programming Language and Libraries**

**Work Carried Out:**



**Prediction of House Pricing Using Machine Learning with Python:**

Data Collection : Data collection is the process of gathering and measuring information on targeted variables in an established system, which then enables one to answer relevant questions and evaluate outcomes. Data collection is a research component in all study fields, including physical and social sciences, humanities, and business.

Data Visualization : Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

Data Pre-Processing : Data preprocessing in Machine Learning refers to the technique of preparing (cleaning and organizing) the raw data to make it suitable for a building and training Machine Learning models.

Data Cleaning : Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabelled.

**Machine Learning based Predicting House Prices using Regression Techniques**

Linear Regression, Support Vector Machine, K-Nearest Neighbors (KNN) and Random Forest Regression and an ensemble approach by combining KNN and Random Forest Technique used for predicting the property's price value.

The ensemble approach predicted the prices with least error of 0.0985 and applying PCA didn't improve the prediction error.

Several studies have also focused on the collection of features and extraction procedures. Wu, Jiao Yang has compared various feature selection and feature extraction algorithms combined with Support Vector Regression.

Some researchers have developed neural network models to predict house prices.

Limsombunchai, compared hedonic pricing structure with artificial neural network model to predict the house prices.  Cebula applies the hedonic price model to predict housing prices in the City of Savannah, Georgia.

Jirong, Mingcang and Liuguangyan apply support vector machine (SVM) regression to predict China's housing prices from 1993 to 2002.

They have applied the genetic algorithm to tune the hyper-parameters in the SVM regression model.   Tay and Ho compared the pricing prediction between regression analysis and artificial neural network in predicting apartment's prices.

## V.  RESULTS AND DISCUSSION

The purpose of this system is to determine the price of a house by looking at the various features which are given as input by the user. These features are given to the ML model and based on how these features affect the label it gives out a prediction. This will be done by first searching for an appropriate dataset that suits the needs of the developer as well as the user. Furthermore, after finalizing the dataset, the dataset will go through the process known as data cleaning where all the data which is not needed will be eliminated and the raw data will be turned into a .csv file.

Moreover, the data will go through data preprocessing where missing data will be handled and if needed label encoding will be done. Moreover, this will go through data transformation where it will be converted into a NumPy array so that it can finally be sent for training the model. While training various machine learning algorithms will be used to train the model their error rate will be extracted and consequently an algorithm and model will be finalized which can yield accurate predictions.

Users and companies will be able to log in and then fill a form about various attributes about their property that they want to predict the price of. Additionally, after a thorough selection of attributes, the form will be submitted. This data entered by the user will then go to the model and within seconds the user will be able to view the predicted price of the property that they put in.

## VI.  CONCLUSION

We anticipate that our advanced machine learning model will provide accurate and insightful predictions of future real estate prices. By leveraging historical data and sophisticated algorithms, we aim to offer users a valuable tool for making informed decisions in the dynamic real estate market. Our hope is that this model will contribute to optimized investment strategies, improved risk management, and enhanced overall decision-making, empowering stakeholders to navigate the real estate landscape with confidence and foresight

## VII.  REFERENCES

[1]  A. Adair, J. Berry, W. McGreal, Hedonic modeling, housing submarkets and residential valuation, Journal of Property Research, 13 (1996) 67-83.

[2]  O. Bin, A prediction comparison of housing sales prices by parametric versus semi-parametric regressions, Journal of Housing Economics, 13 (2004) 68-84.

[3]  T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How many trees in a random forest?" In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7376 LNAI, 2012, pp. 154–168, ISBN: 9783642315367.
DOI: 10.1007/978-3-642-31537-4\13

[4]  J. Schmidhuber, "Multi-column deep neural networks for image classification," in Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), ser. CVPR '12, Washington, DC, USA: IEEE Computer Society, 2012, pp. 3642–3649, ISBN: 978-1-4673-1226-4. [Online].

[5]  T. Kauko, P. Hooimeijer, J. Hakfoort, capturing housing market segmentation: An alternative approach based on neural network modeling, Housing Studies, 17 (2002) 875-894.

[6]  R. J. Shiller, "Understanding recent trends in house prices and home ownership," National Bureau of Economic Research, Working Paper 13553, Oct. 2007. DOI: 10.3386/w13553. [Online].

[7]  The elements of statistical learning, Trevor Hastie - Random Forest Generation.

[8]  S. Yin, S. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," IEEE Transactions on Industrial Electronics, 2014.