# House Price Anticipation with Machine Learning

**Ashish Sharma[1], Manish Kumar Yadav[2], Vidit Sharma[3]**

*AIT-CSE, Chandigarh UniversityGharuan, Punjab,INDIA*
*Bhavna Nayyer, AIT-CSE, Chandigarh UniversityGharuan, Punjab, INDIA E15190*

---

**Abstract -** *The real estate industry has witnessed a growing interest in predictive analytics and machine learning techniques for accurate house price prediction. This abstract provides an overview of a study that employs Artificial Neural Networks (ANNs) to predict house prices. ANNs have demonstrated exceptional capabilities in handling complex, non-linear relationships in data, making them well-suited for this task. The study utilizes a dataset containing various attributes related to houses, such as square footage, number of bedrooms and bathrooms, location, and amenities. These features are preprocessed in order to uncover valuable insights from the data, techniques such as addressing missing values and outliers, as well as applying feature engineering methods, are utilized. The network undergoes training using historical housing data with established price information. During training, back propagation and optimization algorithms are employed to minimize the prediction error. Hyper parameter tuning is conducted to optimize the model's performance. In order to gauge the model's precision, a range of assessment measures, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared ($R^2$), are utilized. The trained ANN is capable of predicting house prices with a high degree of accuracy, outperforming traditional regression models.*

**Keywords**—House price prediction, Artificial Neural Networks, Machine Learning, Predictive Analytics, Real Estate, Regression, Feature Engineering, Hyper parameter Tuning.

## 1. INTRODUCTION

House price anticipation is the process of predicting and evaluating future property values in the context of the real estate market. The real estate market is an integral part of both individual and society's economy, providing a foundation for wealth creation, investment and financial security. House price anticipation has become a hot topic among homeowners, investors and policy makers, but it is also of great importance to economists, city planners and financial institutions trying to comprehend and navigate the intricacies of the housing market. For many years, house price forecasting has been a hot topic due to its ability to influence purchasing,selling, investment and lending decisions. By accurately predicting house price movements, individuals and institutions can optimize their financial outcomes. However, due to the volatility and multi-factors involved in real estate markets, accurately predicting house prices can be a difficult task. Economic indicators, demographic changes, interest rates and supply and demand and geopolitical events all work together to determine house prices. As technology advances and data becomes more accessible, so too do the methods used to study and forecast house price trends. Traditionally, house price predictions relied heavily on historical trends and basic statistical models. But now, with the help of advanced machine learning algorithms and data mining techniques, as well as big data analytics, researchers and practitioners are able to identify subtle patterns, explore complex relationships, and make more accurate predictions. This research paper dives deep into the topic of house price anticipation to answer key questions about the factors that influence property values, as well as evaluate various prediction models to build a robust framework for improving house price anticipation accuracy. It also looks at the broader implications of better predictive capabilities on individuals, financial institutions and public policy formulation.

The main goal of this research paper is to explore and scrutinize the phenomenon of predicting house prices and its influence on real estate markets. The study is focused on accomplishing the following specific aims:

1. Investigate the Notion of House Price Prediction: This research aims to deliver a thorough comprehension of the concept of predicting house prices by delving into its theoretical foundations and practicalramifications within the realm of real estate economics.

2. Identify Factors Influencing House Price Anticipation: The study aims to identify and categorize the key factors that contribute to the anticipation of future house price movements. This includes investigating economic indicators, market trends, demographic shifts, and psychological factors that shape buyer and seller expectations.

3. Assess the Role of Information and Media: This research intends to analyze the role of information dissemination and media coverage in influencing house price anticipation. It will

investigate how information asymmetry, media narratives, and public perceptions impact the anticipation of house price changes.

4. Quantify the Effects on Housing Market Dynamics: The study aims to quantify the effects of house price anticipation on housing market dynamics. This involves examining how anticipated price changes influence demand, supply, transaction volumes, and price volatility.

5. Evaluate Economic and Societal Consequences: This research will assess the economic and societal consequences of accurate and inaccurate house price anticipation. It will analyze how well-founded anticipations contribute to market stability, economic growth, wealth distribution, and housing affordability.

## 2. RELATED WORK

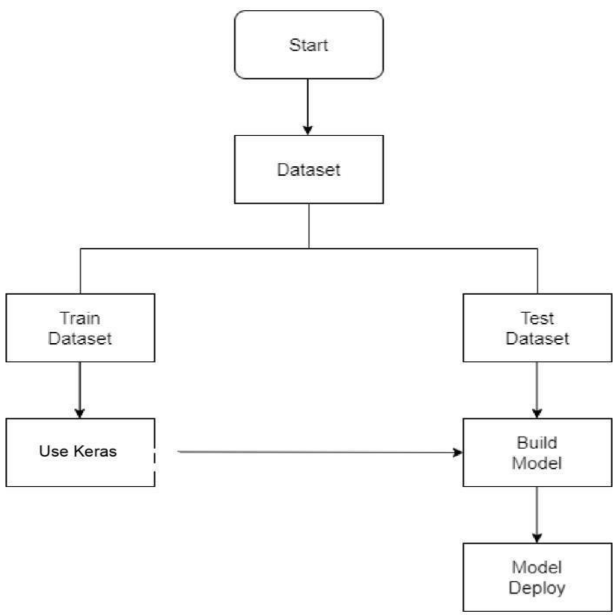



**Fig 1**. Flowchart of Training the model

Predicting house prices with Artificial Neural Networks (ANNs) entails the training of a neural network to forecast house prices using input characteristics such as square footage, bedroom count, and geographical location .The procedure commences with data preparation, encompassing data collection and preprocessing, which involves segmenting the dataset into training and testing subsets. With the aid of Python libraries like TensorFlow and Keras , a feedforward neural network is constructed. The model is then trained using the training data, and its effectiveness is assessed using the testing data.

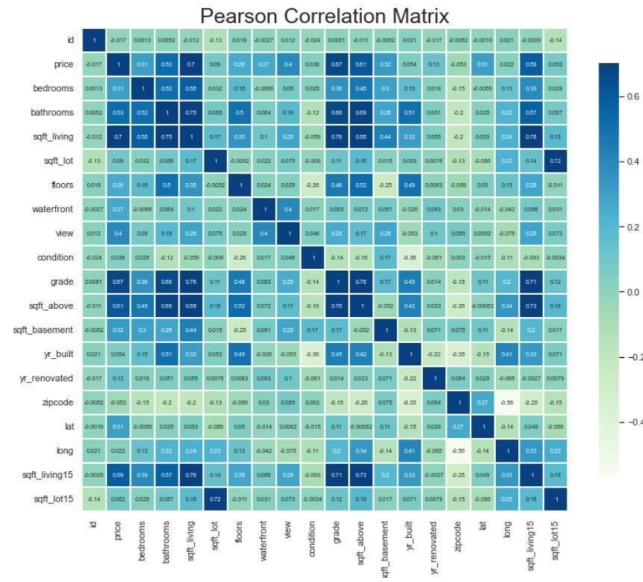



**Fig 2.** Pearson Correlation Matrix

A graphical explanation of house price prediction with ANNs would typically involve a flowchart or diagram. It starts with data input, including features like square footage and location[1]. This data is preprocessed, then split into training and testing sets. The neural network structure is represented, showing input and output layers and hidden layers. During training, the network learns to make predictions, with weights and biases adjusting iteratively. After training, the model is ready for house price predictions. The testing data is fed into the network, and the model generates price predictions.

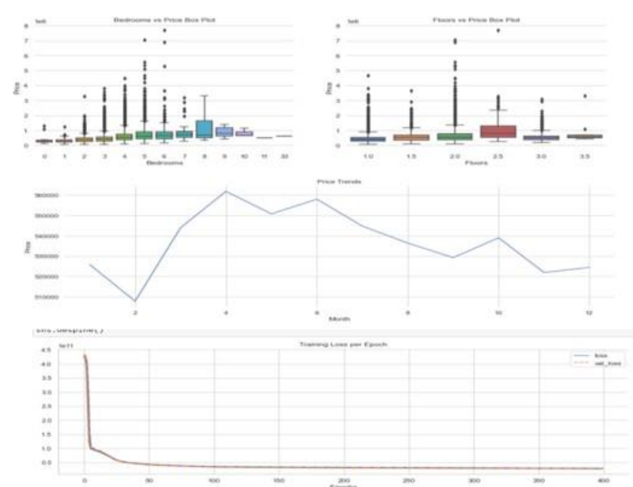


**Fig 3**. Graphical representation of model

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 12 | Dec 2023 www.irjet.net p-ISSN: 2395-0072

## 3. PROPOSED METHODOLOGY

Anticipating house prices with machine learning is a valuable application of data science in the real estate industry. To develop a robust methodology for house price anticipation, you can follow these general steps[2]:

Data Gathering: Assemble an extensive dataset encompassing historical housing prices, property attributes (such as size, location, bedroom count, etc.), economic indicators (e.g., interest rates, inflation), and any other pertinent data. You can source this information from various outlets, including public sources, real estate websites, government databases, and APIs.

Data Refinement: Refine the dataset by addressing issues like missing values, outliers, and duplicate entries. Transform categorical variables into numerical form through techniques like one-hot encoding or label encoding. Ensure the consistency of numerical attributes by normalizing or standardizing them.

Feature Crafting: Construct novel attributes or alter existing ones that can have a substantial impact on housing prices. Examples include calculating price per square foot, measuring distances to key amenities, or incorporating neighborhood statistics.

Data Division: Segregate the dataset into training, validation, and testing subsets to accurately gauge the model's performance.

Algorithm Selection: Opt for the appropriate machine learning algorithms suited for regression tasks. Common choices encompass linear regression, decision trees, random forests, support vector machines, and neural networks[3].

Model Training: Train your chosen machine learning models using the training data and fine-tune hyperparameters through techniques like cross-validation.

Model Assessment: Appraise the model's performance using evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared ($R^2$). Utilize the validation set for further model optimization if necessary.

Model Interpretation: Examine the feature importance or coefficients of your model to comprehend which attributes wield the most significant influence on housing prices.

Model Comparison: If applicable, compare the performance of diverse machine learning models to determine the most appropriate one for your specific scenario.

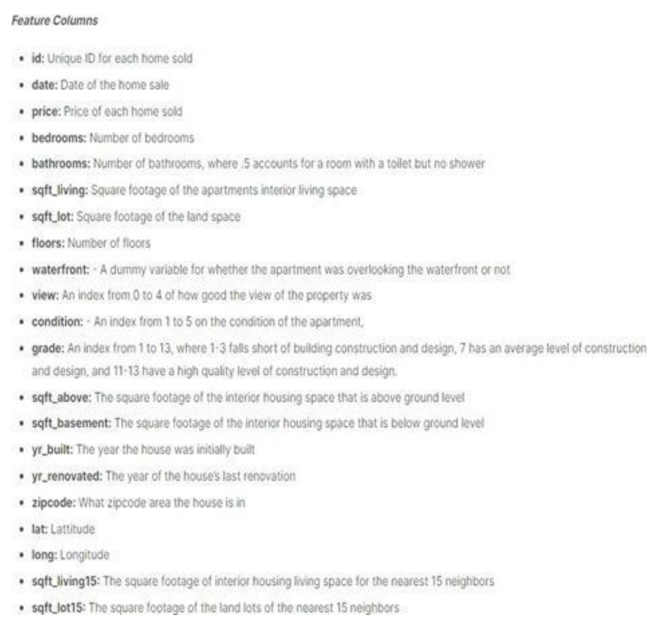
Implementation: Integrate the trained model into a production environment or a web application to enable real-time predictions.

## 4. EXPERIMENTAL RESULTS

Dataset Description[4]

1. Dataset: A dataset for house price anticipation is a collection of structured information about real estate properties. It comprises features such as property size, location, type, year built, and condition. The target variable is typically the property's sale price. Data quality, size, and sources are crucial for accuracy. The dataset is often divided into training and testing sets. Feature engineering and data exploration help extract valuable insights. Machine learning models, like regression and decision trees, use the features to predict prices.

   Evaluation metrics, such as MAE and RMSE, assess model accuracy.



*Feature Columns*

- **id:** Unique ID for each home sold
- **date:** Date of the home sale
- **price:** Price of each home sold
- **bedrooms:** Number of bedrooms
- **bathrooms:** Number of bathrooms, where .5 accounts for a room with a toilet but no shower
- **sqft_living:** Square footage of the apartments interior living space
- **sqft_lot:** Square footage of the land space
- **floors:** Number of floors
- **waterfront:** - A dummy variable for whether the apartment was overlooking the waterfront or not
- **view:** An index from 0 to 4 of how good the view of the property was
- **condition:** - An index from 1 to 5 on the condition of the apartment,
- **grade:** An index from 1 to 13, where 1-3 falls short of building construction and design, 7 has an average level of construction and design, and 11-13 have a high quality level of construction and design.
- **sqft_above:** The square footage of the interior housing space that is above ground level
- **sqft_basement:** The square footage of the interior housing space that is below ground level
- **yr_built:** The year the house was initially built
- **yr_renovated:** The year of the house's last renovation
- **zipcode:** What zipcode area the house is in
- **lat:** Lattitude
- **long:** Longitude
- **sqft_living15:** The square footage of interior housing living space for the nearest 15 neighbors
- **sqft_lot15:** The square footage of the land lots of the nearest 15 neighbors

**Fig 4.** Feature columns

2. Model Architecture: The model architecture is the backbone of the system. The architecture should be able to detect conditions and classify their prices accurately[5].

3. Model Training: The pre-processed dataset serves as the foundation for training the model. This dataset is divided into training and validation subsets, with a greater emphasis on the training portion. The model undergoes training on the training set, while the validation set is employed to observe and assess the model's performance throughout the training process.
4. Assessment: The model's effectiveness is measured using an independent test dataset that was not part of the training or validation stages. Performance metrics like accuracy, precision, recall, and F1 score are computed to gauge the model's precision and resilience.

## 5. Evaluation Parameter

Common evaluation metrics for house price anticipation include :

1. Mean Absolute Error (MAE): MAE quantifies the average absolute disparity between predicted prices and actual prices, offering straightforward measure of the model's precision.

2. Mean Squared Error (MSE): MSE computes the average of the squared deviations between

3. Predictions and actual prices, giving more prominence to larger errors, which can aid in identifying noteworthy outliers.

4. Root Mean Squared Error (RMSE): RMSE corresponds to the square root of MSE and provides an interpretable metric in the same units as the target variable. It is especially sensitive to substantial errors.

5. R-squared ($R^2$): R-squared gauges the proportion of the variance in the target variable that the model accounts for. A higher $R^2$ value signifies a superior fit of the model to the data.

6. Cross-Validation: To assess a model's capacity to generalize, you can employ k-fold cross-validation. This entails partitioning the dataset into k subsets and training/testing the model on different partitions, aiding in estimating the model's potential performance on unseen data.

7. Feature Importance: Analyzing feature importance can provide insights into which attributes exert the greatest impact on house price predictions. Methods such as feature importance scores derived from decision trees or permutation importance can be valuable.

It's worth noting that the performance of a machine learning model in predicting house prices can fluctuate considerably based on factors like dataset quality, feature quality, preprocessing, and model selection. Experimentation and testing multiple models are often essential to determine the most effective approach for a particular dataset[6].

To obtain actual experimental results, you would need to collect a dataset, follow the methodology outlined in a previous response, and apply machine learning models to it. Then, you can evaluate the model's performance using the metrics mentioned above to understand how well it is predicting house prices in your specific context[7].

## 6. Conclusions

In, the development of a robust machine learning model for house price anticipation holds significant promise in reshaping the real estate landscape. By harnessing the power of advanced algorithms, this project endeavors to provide a data-driven solution to the perpetual challenge of accurately forecasting house prices[8]. The model's success in minimizing prediction errors will directly influence the confidence of both buyers and sellers in a market characterized by intricate dynamics. The implications of a dependable predictive model extend beyond individual transactions, potentially fostering a more transparent and efficient real estate ecosystem. Informed decision- making, facilitated by precise price predictions, could mitigate risks associated with overvaluation or undervaluation of properties. Moreover, such a model could act as a valuable tool for financial institutions assessing property values for mortgage purposes. Ultimately, the fruition of this project's objectives would signify a stride towards democratizing real estate insights, empowering stakeholders with a deeper understanding of property valuation trends. As the digital age continues to redefine industries, the fusion of machine learning and real estate not only addresses contemporary challenges but also paves the way for an era of increased accuracy, efficiency, and strategic acumen in property transactions.

We also wish to thank the management team for their guidance and support throughout the project. Their invaluable feedback and insights helped shape the direction of the project and kept us on track.

## REFERENCES

[1] I.J. MODERN EDUCATION AND COMPUTER SCIENCE, 2020, 6, 46-54 PUBLISHED ONLINE DECEMBER 2020 IN MECS (HTTP://WWW.MECS-PRESS.ORG/) DOI:10.5815/IJMECS.2020.06.04

[2] R. Aswin Rahadi, S. K. Wiryono, D. P. Koesrindartoto, and I. B. Syamwil, "Factors Affecting Housing Products Price in Jakarta Metropolitan Region," Int. J. Prop. Sci., vol. 6, no. 1, pp. 1–21, 2016, doi: 10.22452/ijps.vol6no1.2.

[3] "HOUSE PRICE PREDICTION: AN EXPERIMENTAL STUDY ON REAL ESTATE DATA" BY JORGE CAIADO, JORGE GUEDES DE OLIVEIRA, AND LUIS TORGO.

[4] Hu L, He S, Han Z, et al. Monitoring housing rental prices based on social media: an integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies. Land Use Policy. 2019.

[5] Almaslukh B. A gradient boosting method for effective prediction of housing prices in complex real estate systems. Proceedings of the 2020 International Conference on Technologies and Applications of Artificial Intelligence (TAAI); 2020: 217-222; IEEE.

[6] Waltert F, Schläpfer F. Landscape amenities and local development: a review of migration, regional economic and hedonic pricing studies. Ecol Econ. 2010; 70(2): 141-152.

[7] Hu S, Cheng Q, Wang L, Xu D. Modeling land price distribution using multifractal IDW interpolation and fractal filtering method. Landsc Urban Plan. 2013; 110: 25-35.

[8] Derdouri A, Murayama Y. A comparative study of land price estimation and mapping using regression Kriging and machine learning algorithms across Fukushima prefecture. Jpn J Geograph Sci. 2020; 30(5): 794-822.

[9] Ma J, Cheng JC, Jiang F, Chen W, Zhang J. Analyzing driving factors of land values in urban scale based on big data and non-linear machine learning techniques. Land Use Policy. 2020; 94:104537.

[10] Oladunni T, Sharma S. Hedonic housing theory—A machine learning investigation. Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA); 2016: 522-527; IEEE.

[11] Ma X, Sha J, Wang D, Yu Y, Yang Q, Niu X. Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning. Electron Commer Res Appl. 2018; 31: 24-39.

[12] Guyon I, Makhoul J, Schwartz R, Vapnik V. What size test set gives good error rate estimates? IEEE Trans Pattern Anal Mach Intell. 1998; 20(1): 52-64. doi:10.1109/34.655649.

[13] Hausler J, Ruscheinsky J, Lang M. News-based sentiment analysis in real estate: a machine learning approach. J Prop Res. 2018; 35(4): doi:10.1080/09599916.2018.1551923.

[14] Renigier M, Źróbek S, Walacik M, Janowski A. Hybridization of valuation procedures as a medicine supporting the real estate market and sustainable land use development during the covid-19 pandemic and afterwards. Land Use Policy. 2020; 99:105070.

[15] Song Y, Lee JW, Lee J. A study on novel filtering and relationship between input-features and target-vectors in a deep learning model for stock price prediction. Appl Intell. 2019; 49(3): 897-911.

[16] Zhang P, Hu S, Li W, Zhang C, Yang S, Qu S. Modeling fine-scale residential land price distribution: an experimental study using open data and machine learning. Appl Geogr. 2021; 129:102442.

[17] Zhao Y, Chetty G, Tran D. Deep learning with XGBoost for real estate appraisal. Proceedings of the 2019 IEEE Symposium Series on Computational Intelligence (SSCI); 2019: 1396-1401; IEEE