



House Price Prediction Analysis using Machine Learning

Aniket Singh¹, Adarsh Kumar Singh², Aditya Raj³, Harshit Jain⁴, Mrs. Asha M S⁵

¹²³⁴Student [BE], Department of Computer Science and Engineering, Dayananda Sagar Academy of Technology and Management (DSATM), Bengaluru, Karnataka, India.

⁵Assistant Professor, Department of Computer Science and Engineering, Dayananda Sagar Academy of Technology and Management (DSATM), Bengaluru, Karnataka, India.

Abstract— House price prediction is a popular topic, and research teams are increasingly performing related studies by using deep learning or machine learning models. However, because some studies have not considered comprehensive information that affects house prices, prediction results are not always sufficiently precise. Therefore, we propose an end to end joint self-attention model for house prediction. In this model, we import data on public facilities such as parks, schools, and mass rapid transit stations to represent the availability of amenities, and we use satellite maps to analyze the environment surrounding houses. We adopt attention mechanisms, which are widely used in image, speech, and translation tasks, to identify crucial features that are considered by prospective house buyers. The model can automatically assign weights when given transaction data. Our proposed model differs from self-attention models because it considers the interaction between two different features to learn the complicated relationship between features in order to increase prediction precision. We conduct experiments to demonstrate the performance of the model. Experimental data include actual selling prices in real estate transaction data for the period from 2017 to 2018, public facility data acquired from the Taipei and New Taipei governments, and satellite maps crawled using the Google Maps application programming interface.

Keywords—Linear Regression, Machine Learning, Histogram

I. INTRODUCTION

House price prediction is currently a hot topic. The purpose of house price prediction is to provide a basis for pricing between buyers and sellers. By viewing transaction records, buyers can understand whether they have received a fair price for a house, and sellers can evaluate the price at which they can sell a house along a specific road section. House price prediction

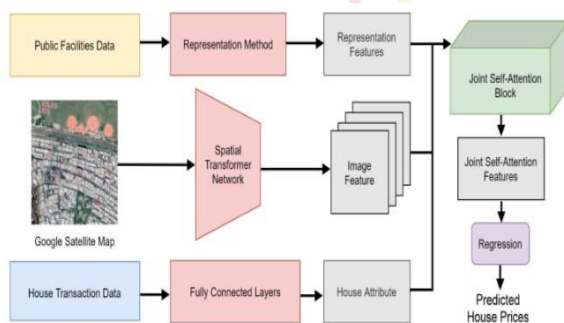
is also used in financial technology applications, which require a reasonable evaluation system for mortgage calculation and house auctions. Studies on house price prediction have been conducted in the United States and Europe and most of these studies have adopted machine learning-based methods to predict real estate prices. Related research has also been established even in developing countries such as Brazil. Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis of the objectives of the present study is to use actual selling prices in real estate transaction data to perform predictions. After reviewing research from these countries, we identify several factors that affect house prices and categorize such factors into two groups. The first group comprises house conditions, such as age, building material, building area, and parking space; the second group comprises conditions of the surrounding environment, including public facilities, transportation facilities, and educational institutions. Increasing numbers of studies on house price estimation have adopted machine learning and deep learning methods; researchers have mostly been concerned with the accuracy of forecasting results. However, some studies have considered only house attributes, and they have not accounted for information on surrounding environments. Therefore, using the Google Maps application programming interface (API), we crawl satellite maps of houses and surrounding public facilities as heterogeneous input data for the prediction model. We use house transaction data, the satellite maps, and public facility data to experimentally test our hypothesis. To process the satellite maps, we adopt a spatial transformer network to extract image features. An STN has rotation invariant properties, meaning that image extraction features do not change markedly when an image is rotated by a specific angle.

Different groups of buyers may focus on different house attributes. For example, nuclear families may focus on parks and nearby schools for their children. Therefore, if a house's crucial features identified through an attention mechanism match the needs of a prospective buyer, then the house can be recommended to the buyer. Accordingly, our final research goal is to improve prediction accuracy and investigate features influencing pricing results. We employ an attention mechanism and import heterogeneous data into our developed attention model. These data and the introduced methods not only improve house price prediction performance but also reveal factors influencing house prices. In addition to achieving accurate house price prediction, we identify the features contributing to house prices. Therefore, we develop an attention-based model that can learn the interaction between features and summarize crucial attributes for house buyers. Additionally, we design novel methods to effectively extract features from heterogeneous data to help improve model performance.

II. LITERATURE SURVEY

House is one of human life's most essential needs, along with other fundamental needs such as food, water, and much more. Demand for houses grew rapidly over the years as people's living standards improved. While there are people who make their house as an investment and property, yet most people around the world are buying a house as their shelter or as their livelihood. According to [1], housing markets have a positive impact on a country's currency, which is an important national economy scale. Homeowners will purchase goods such as furniture and household equipment for their home, and homebuilders or contractors will purchase raw material to build houses to satisfy house demand, which is an indication of the economic wave effect created by the new house supply. Besides that, consumers have capital to make a large

investment, and the construction industry is in good condition can be seen through a country's high level of house supply. According to numerous international organizations and human rights have emphasized house importance. House is profoundly rooted in the economic, financial, and political structure of each country. Nevertheless, reported that the fluctuation of house prices has always been an issue for house owners, buildings and real estate, besides stated that house has become unaffordable as there is substantial price growth in several countries in the House Price. Residents' quality of life as well as national economy depends on the potential house price increase. Ultimately, this issue will affect investors who are making their house as an investment. An increase in house demand occurs each year, indirectly causing house price increases every year. The problem arises when there are numerous variables such as location and property demand that may influence the house price, thus most stakeholders including buyers and developers, house builders and the real estate industry would like to know the exact attributes or the accurate factors influencing the house price to help investors make decisions and help house builders set the house price. House price prediction can be done by using a multiple prediction models (Machine Learning Model) such as support vector regression, artificial neural network, and more. There are many benefits that home buyers, property investors, and house builders can reap from the house-price model. This model will provide a lot of information and knowledge to home buyers, property investors and house builders, such as the valuation of house prices in the present market, which will help them determine house prices. Meanwhile, this model can help potential buyers decide the characteristics of a house they want according to their budget. Previous studies focused on analyzing the attributes that affect house price and predicting house price based on the model of machine learning separately. However, this article combines such a both predicting house price and attributes together. In this article, literature review focuses on predicting house price based on the model of machine learning as well as analyzing attributes primarily used in previous study that affect house price. This paper was arranged as follows: the first section summarizing overall of this study. Second section described the common attributes used in prediction of house price around the world. It was followed by a brief discussion of machine learning model used in previous study to predict house price.



investment, and the construction industry is in good condition can be seen through a country's high level of house supply. According to numerous international organizations and human rights have emphasized house importance. House is profoundly rooted in the economic, financial, and political structure of each country. Nevertheless, reported that the fluctuation of house prices has always been an issue for house owners, buildings and real estate, besides stated that house has become unaffordable as there is substantial price growth

III. PROPOSED SYSTEM

The problem is creating the hypothesis function that may give the prediction of the target value based on the data given as the training part.

Then see or analyze the prediction on the testing part of the data. Here the data given is on the house price and its respective features which accommodate the price of the house. Thus to build the machine to learn the data features and predict the price accurate is the challenging task.

This will also help the society of the real estate builder to easily predict the price of the land, house etc according to their feature with the help of this model. The data set for this thesis is taken from Kaggle's Housing Data Set. Data set is simple and this thesis aims at the prediction of the house price (residential) in Bangalore, India. Thus the data has been downloaded from the Kaggle Housing Datasets.

The fetching operation is done by the help of the pandas library function as in the format of .csv file and giving the path where data is stored.

After fetching the data, some cleaning process is applied to the data to make it provide useful information. Thus the missing values in the attribute is checked and clean it out i.e. drop attribute if it is not much useful feature or fill the missing value by taking the median of the all values.

Then the further process applying machine learning is processed and three regression technique is applied to the dataset linear regression, random forest.

IV. METHODOLOGY

According to the paradigm of evaluating the house demand can be classified into two classes which are the traditional method and the advanced valuation method. The traditional valuation scheme, including multiple regression method and stepwise regression process, whilst hedonic pricing tool, artificial neural network (ANN) and spatial analysis framework are advanced valuation method. The model selection to be used to predict house price is quite critical as varieties of models are available. One of the most commonly utilized models in this research field is Regression Analysis which is used in many studies, including Another common model for house price predictions is the Support Vector Regression (SVR) A Regression Analysis Hedonic Price Model The housing market is slightly different from normal good consumption. According to housing market is unique because it displays the characteristics of resilience, flexibility and spatial fixity. Therefore, hedonic approach is preferred to accurately predict market differential. conceived hedonic model back in 1939, but this research was popularized in the early 1960s with comprehensive uses by Zvi Griliches and Rosen. In the early 1930s, Court used this model to analyse automotive value in pricing and quality characteristics defined hedonic as "the implicit prices of attributes and are revealed to economic agents from the observed prices of differentiated products and the specific quantities of their characteristics." Following

years of progress, Rosen applied the approach to the residential home price study and became commonly included in real estate sector research. Rosen's philosophy or model comprises two separate phases. The regression of a product price on its attributes is performed in the initial stage to calculate the aggregate price of the component. A measure of a goods price will be determined in the first stage, but the inverse demand function cannot be generated at this stage. Thus, the second stage of estimation is needed to identify the inverse demand function that can be derived from the first stage implicit price function. In an earlier study, a study compared three commonly used house price measurement methods which are simple average method, hedonic model, and matching approach. The result found that when adopted on the housing market, two methods that are simple average method and matching approach were proven biased. Thus, the hedonic model provides the highest results relative to those two most commonly encountered versions. Hedonic pricing model is a statistical model that believes the worth of the property is the sum of all its attributes based on hedonic market theory. ii. Multiple Linear Regression analysis is a model used to determine the relationship between variables. In order to evaluate the correlation of the variables, the correlation coefficient or regression equation can be used. Multiple regression models can determine which characteristics are the most important to explain the dependent variable. Multiple regression analysis also allows certain price predictions by capturing independent and dependent variable data. In the power of the multiple regression model can be seen when the value of the relationship between dependent and independent variables is measured use multiple regression modelling to describe improvements to an independent variable with a dependent variable. This model can be achieved using the house price projection as separate and dependent variables like house prices, house size, property sort, number of bedrooms, and many more. Therefore, the house price is set as a target or dependency variable, while other attributes are set as independent variables to determine the main variables by identifying the correlation coefficient of each attribute. B. Support Vector Regression Support vector regression is a predictive model based on SVM, a neural network that usually has three layers, a powerful form of supervised learning. The model is based on a subset of training data. The advantages of support vector regression are that it is capable of processing non-linear results, provides only one possible optimal solution, and able to overcome a small sample learning issues. The potential to produce market predictions in several markets, including real estate, shows that this model can overcome the non-linear regression problems and small sample learning problems. Moreover, as this model did not depend on probability distribution assumptions, and the ability of mapping the input attribute, either linear or

non-linear, this model was commonly used at house price modelling. Support vector regression offers huge benefits in so many aspects as this model can avoid over-fitting problems, while ensuring a single optimum solution by minimizing structural risks and empirical risks. In this field of study, support vector regression is used to collect details on neighborhood, structural and locational attributes. C. Artificial neural Network In 1958, created artificial neural network known as ANN. Walter Pitts and Warren McCulloch published a paper entitled "A Logical Calculus of Ideas Immanent in Nervous Activity" in the year 1943 which notes that a neural network may artificially be created, based on the role and structure of a biological neural network. In another research, as this model would often promote learning, artificial neural networks are claimed to be artificial brain diagrams. The artificial neural network model has always been selected when a non-linear attribute is involved. The analysis of home price estimation should also use this model as a spatial consideration for the price of housing is also non-linear. Therefore, as in their study produces a good result, thus it is promising to provide an exact predictive model utilizing the artificial neural network algorithm. This system, however, has very limited performance. ANN can model complex non-linear relationships as house price predictions involve many non-linear variables. D. Gradient Boost Gradient boosting was created by in 1999 and is a commonly used machine learning algorithm because of its performance, consistency and interpretability. Gradient boosting delivers state-of-the-art in various machine learning activities, such as multistage classification, click prediction and ranking. With the advent of big data in recent years, gradient boosting faces new challenges, especially with regard to the balance between accuracy and performance. There are few parameters for gradient boosting. To ensure a dynamic balance between fit and regularity, the following steps can be taken to select parameters: (1) Setting regularization parameters (λ , α), (2) reducing learning rate and decide those optimal parameters again.

CONCLUSION

This paper examined and analyzed the current research on the significant attributes of house price and analyzed the data mining techniques used to predict house price. Technically, houses with a strategic location such as the accessibility to shopping mall or other facilities tend to be more expensive than houses in rural areas with limited numbers of facilities. The accurate prediction model would allow investors or house buyers to determine the realistic price of a house as well as the house developers to decide the affordable house price. This paper addressed the attributes used by previous researchers to forecast a

house price using various prediction models. Taken together, the results of the survey have shown the potential of SVR, ANN and XG Boost in predicting house prices. These models were developed based on several input attributes and they work significantly positive with house price. In conclusion, the impact of this research was intended to help and assist other researchers in developing a real model which can easily and accurately predict house prices. Further work on a real model needs to be done with the utilization of our findings to confirm them.

ACKNOWLEDGMENT

We would like to express our utmost gratitude to our principal Dr. Ravishankar M sir and our vice principal and HOD of CSE dept Dr. C Nandini ma'am for giving us an opportunity to showcase our project and research work and our beloved guide Mrs. Asha MS ma'am for guiding us in innumerable ways towards the development and success of the project.

V. REFERENCES

1. Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along with Joint Self-Attention Mechanism. - IEEE APRIL 2021 (PEI-YING WANG, CHIAO-TING CHEN, JAIN-WUN SU, TING-YUN WANG, AND SZU-HAO HUANG)
2. Housing Price Prediction via Improved Machine Learning Techniques. 2019 (Quang Truong, Minh Nguyen, Hy Dang, Bo Mei)
3. Real estate price prediction with regression and classification CS 229 Autumn, 2016 (Hujia Yu, Jiafu Wu)
4. House Price Prediction using a Machine Learning Model : A survey of Literature – International general of Modern Education and Computer sciences Dec 2020 (Nor Hamizah Zulkifley, Shuzlina Abdul Rahman, Nor Hasbiah Ubaidullah, Ismail Ibrahim)
5. M. Risse and M. Kern, "Forecasting house-price growth in the euro area with dynamic model averaging," North Amer. J. Econ. Finance, vol. 38, pp. 70–85, Nov. 2016.
6. B. Afonso, L. Melo, W. Oliveira, S. Sousa, and L. Berton, "Housing prices prediction with a deep learning and random forest ensemble," in Proc. Anais do 16th Encontro Nacional de Inteligência Artif. e Computacional, 2019.

7. M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in Proc. Adv. Neural Inf. Process. Syst., 2015, pp. 2017–2025.

8. Z. Peng, Q. Huang, and Y. Han, "Model research on forecast of secondhand house price in Chengdu based on XGboost algorithm," in Proc. IEEE 11th Int. Conf. Adv. Infocomm Technol. (ICAIT), Oct. 2019, pp. 168–172.

