Machine learning based approach for predicting house price in real estate FREE

Mohammed Ali Shaik ✉ ; P. Praveen; T. Sampath Kumar; Masrath Parveen; Swetha Mucha

Check for updates

View Online

Export Citation

07 June 2024 08:52:49

AIP Conference Proceedings

AIP Publishing

Machine Learning Based Approach for Predicting House Price in Real Estate

Mohammed Ali Shaik[1, a)], P Praveen[1], T Sampath Kumar[1], Masrath Parveen[2] and Swetha Mucha[3]

[1]*School of Computer Science and Artificial Intelligence, SR University, Warangal, Telangana, India.*
[2]*Vidya Jyothi Institute of Technology, Hyderabad, India.*
[3] *Sumathi Reddy Institute of Technology for Women, Warangal, Telangana, India*

a) Corresponding author: niharali@gmail.com

Abstract. The proposed application must assist a targeted user to obtaining real estate property information with elegant correctness and precise output. By implementing the proposed algorithmic prediction process, which will forecast various property prices based on several criterions. We're going to use datasets from the landmark we're going to forecast, and this landmark can be changed according to the area we want to predict. We'll use a linear regression technique to predict, and by utilizing Flask framework more accurate results can be acquired. We discovered that by training data to the maximum, we were able to achieve absolute outcomes. We can forecast working in this subject by leveraging technologies such as Python, which allows us to train and collect more entertaining data. In comparison to other investments, real estate property prices are unaffordable. We can say that we are able to predict.

INTRODUCTION

Purchasing a home is a significant choice in a person's life that takes considerable consideration and research. One would like to purchase a home at the greatest possible price with little risk, and for it to be an excellent long-term investment. Home purchasers are guided by online portals, real estate agents, and realtors who allow them to compare different residences available for purchase. Because our concept is entirely based on the prediction of raw real estate data that we have gathered from past data.

There are many people who wish to acquire properties or at least consider buying them; nevertheless, it is not ideal for any buyer to buy a home without first inspecting or comparing it to others. Private investigations are preferred before purchasing any property, which is where our initiatives come in. Using machine learning algorithms such as Linear Regression, we are attempting to forecast property values. Our research attempts to identify the greatest regions for a real estate developer to invest in, as well as individual purchasers seeking for a location to create a new apartment building or acquire. Another objective is to forecast property prices in a given county over the following few years.

REVIEW OF LITERATURE

"Machine learning (ML" [1] is a type of "Artificial Intelligence (AI)" that skillfully prepares accessible machines to be trained without being programmed correctly. As ML approach is interested in expanding the number of computer programmes that can be changed when new data becomes insecure and more over ML based algorithms can be divided into three groups: "supervised learning, unsupervised learning, and reinforcement learning".

"Supervised Learning (SL)" [2] is a type of education in which we teach or train a machine utilising properly labelled data, which implies part of the data has already been tagged with the correct response as the machine will

07 June 2024 08:52:49

further provide the new set of examples which are required to analyze the training data and produce the proper outputs from the labeled data using the SL method which is a unhelpful learning process that occurs when a machine is trained on data that is neither classified nor labeled [3], and the algorithm is left to act on such data without supervision. Without having any sorts of prior data training each of the machine's role is to arrange unresolved information according to similarity, pattern, and difference and more over in supervised teaching technique, there exists no teacher who is offered, which means the machine will not be trained. We limit the machine's ability to track down the secret construction in unlabeled data in this way.

ML includes the field of reinforcement learning [4,32,34,45]. It's all about acquiring the right steps to get the most out of a given scenario as most of the applications tend to become robots when used it to determine the best potential action or path in a given situation. "Reinforcement learning (RL)" approach [5] differs from supervised learning in that in SL approach where the training data contains the answer key and thus the model is trained with the correct answer and similarly in RL approach there exists no answer and the reinforcement agent decides what to do inorder to complete any specific task which directly intends to benefit through its experience even if it does not have a preparatory dataset [38,39,40].

One of the mostly utilized ML approaches [6] will predict the application real estate where the real estate market comprises of most competitive aspects in dealing with the terms of pricing which tends to fluctuate most widely as it depends on a variety of criteria that is similar to anticipating property prices that supports the budget allocation and for performing judgments towards both buyers and investors while searching an appropriate home. This is a crucial module [7,25,26] as the results, one of the primary areas of applying machine learning principles to optimize and anticipate prices with high accuracy is locating strategies and deciding appropriate policies. Land price trend studies have long been seen as critical in guiding urban development decisions. The land interaction is a shaky stochastic one. Investors make judgments based on market movements in order to maximize their profits.

As per the research [8,20,23] by implementing simple linear regression and multiple linear regression approaches to obtain the simple least squares for investigating the correlations between different household properties that acquires price as a input field over the residential property. Simple linear regression process includes: "land parcel size, number of bedrooms, year of construction, and other explanatory variables" that multiplies "linear regression" included "land parcel size, number of bedrooms, year of construction, and other explanatory variables".

"Multi-linear regression (MLR)" [9,43,44] findings were found to be biassed as a result of missing of essential parameters in simple linear regression. The most essential element in influencing the price of a residential property was discovered to be home square footage, whereas garage capacity was revealed to be the least relevant component. As per the research [10,12,41,42] which has discovered that in prior regression models, all five variables have significant effects on the dependent variable value if the significance level is set to 0.05. ("floor, heating system, earthquake area, rent value, and land value"). The most significant influences on the price of housing[15] are land value and rental value. They are followed by existing flooring, heating systems, and seismic zones where the other variables were determined to be insignificant in this study, though this could alter depending on sample size. If the sample size grows larger, the regression model is recommended for further research. Indicates the strategic application of multiple regression analysis to a home data set or model's variance in house price.

PROPOSED MODEL

Our primary concept is to design an application that will assist users in locating their selected home. Machine learning and data science are being used to attain this goal. The steps are as follows:

Step 1: To obtain the raw data for kaggle. Kaggle is a data science and machine learning platform that offers the most datasets and tools.

Step 2: We obtained a dataset from Kaggle and began analyzing it. We employed a variety of tools for analysis, including Jupyter Notebook and libraries such as Numpy, Pandas, scikit-learn, and others. We separated useful and unusual data using tools and frameworks.

Step3:¬ Following that, we divided the dataset into training and testing datasets. Following that, we utilised many algorithms such as linear regression, decision trees, and the Lasso (least verify the accuracy & from all algorithms that we checked, absolute shrinkage and selection operator) algorithm. We utilised linear regression for this approach.

Step 4: We then utilized a linear regression technique to predict the price of that real estate using actual data. Machine learning is being used to predict real estate prices. The website that we designed is currently being used by our staff. We used technologies like bootstrap, CSS, and HTML for the frontend. We used Flask technology for the backend.

CONCLUSION

Home purchasers are guided by online portals, real estate agents, and realtors who allow them to compare different residences available for purchase. Because our concept is entirely based on the prediction of raw real estate data that we have gathered from past data, Our research attempts to identify the greatest regions for a real estate developer to invest in, as well as individual purchasers seeking for a location to create a new apartment building or acquire. Another objective is to forecast property prices in a given county over the following few years. Machine learning is a type of artificial intelligence that skillfully prepares accessible machines to be trained without being properly programmed. One of the many uses of machine learning is the prediction of application real estate.

The real estate market is the most competitive in terms of pricing, and it fluctuates widely depending on a variety of criteria, such as anticipating property prices to support budget allocation, and making judgments for both buyers and investors in the search for a home.

As a result, one of the primary areas of applying machine learning principles to optimize and anticipate prices with high accuracy is locating strategies and deciding appropriate policies. The regression model is recommended for future study if the sample size grows. The simple least squares to investigate correlations between various household characteristics and the asking price of a residential property. Other variables in this study were not found to be significant, however this could alter depending on sample size.

REFERENCES

1. A. K. Agogino and K. Tumer, Ensemble clustering with voting active labels, Pattern Recognition Letters, 29(14), 1947–1953, (2008).
2. Babenko and V. Lempitsky, "Tree Quantization for Large-Scale Similarity Search and Classification," in (CVPR, 2015), pp. 651–654.
3. Brandt, "Transform coding for fast approximate nearest neighbor search in high dimensions," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR, 2010), pp. 1815–1822.
4. C. Silpa-Anan and R. Hartley, "Optimised KD -trees for fast image descriptor matching", IEEE Conference on Computer Vision and Pattern Recognition (CVPR, 2008), pp. 1–8.
5. J. Wang, T. Zhang, J. Song, N. Sebe, and H. T. Shen, A Survey on Learning to Hash, IEEE Transactions on Pattern Analysis and Machine Intelligence, 13, 9,123-129, (2017).
6. M. Laszlo and S. Mukherjee, Minimum Spanning Tree Partitioning Algorithm for Micro aggregation, IEEE Trans. on Knowledge and Data Engg., 17(7), 902-911, (2005).
7. Mohamed A. Mahfouz, d M. A. Ismail, Fuzzy Relatives of the CLARANS Algorithm With Application to Text Clustering, International Journal of Electrical and Computer Engineering, 370-377, (2009).
8. Balasundaram A., Kothandaraman D., Ashokkumar S., Sudarshan E. Chest X-ray image based COVID prediction using machine learning 2022 AIP Conference Proceedings 2418 20079 10.1063/5.0081952.
9. Mohammed Ali Shaik, T. Sampath Kumar, P. Praveen, and R. Vijayaprakash, Research on Multi-Agent Experiment in Clustering, International Journal of RecentTechnology and Engineering (IJRTE), 8, 1S4, 1126-1129, (2019).
10. Praveen. P and Ch. JayanthBabu. "Big Data Clustering: Applying Conventional Data Mining Techniques in Big Data Environment.". Innovations in Computer Science and Engineering, Lecture Notes in Networks and Systems 74, Springer Singapore (2019).
11. Yerrolla Chanti, Seena Naik Korra, Bura Vijay Kumar, A Harshavardhan, D Kothandaraman, New Technique using an IoT Robot to Oversight the Smart Domestic Surroundings, Studia Rosenthaliana (Journal for the Study of Research), issue. 0039-3347, 2019.
12. Balasundaram A., Kothandaraman D., Prashanth B., Ashokkumar S. Predicting different facets in the game of cricket using machine learning 2022 AIP Conference Proceedings 2418 20027 10.1063/5.0081951.
13. Mohammed Ali Shaik, Dhanraj Verma, , Agent-MB-DivClues: Multi Agent Mean based Divisive Clustering, Ilkogretim Online - Elementary Education, 20(5), 5597-5603, (2021)
14. T. Sampath Kumar, B. Manula, A Comprehensive Study on Single Sign on Technique, International Journal of Advanced Science and Technology (IJAST) 127, 430-435, (2019).
15. Kumaraswamy E. Key challenges in the diagnosis of cancer using artificial intelligence methods 2022 AIP Conference Proceedings 2418 30049 10.1063/5.0081712.

16. Mohammed Ali Shaik and Dhanraj Verma, "Enhanced ANN training model to smooth and time series forecast", in IOP Conf. Ser.: (Mater. Sci. Eng 2020). 981022038.

17. T. Sampath Kumar, B. Manula,"Competent multi-level encryption methods for implementing cloud security", in IOP Conf. Series: (Materials Science and Engineering, 2020) 022039 IOP Publishing.

18. P Pramod Kumar and K Sagar, "A Relative Survey on Handover Techniques in Mobility Management", in IOP Conf. Ser.: (Mater. Sci 2019), Pp. 12023-12027.

19. Mohammed Ali Shaik, Dhanraj Verma, P Praveen, K Ranganath and Bonthala Prabhanjan Yadav, , RNN based prediction of spatiotemporal data mining, in IOP Conf. Ser.: (Materials Science and Engineering, 2020) 981 022027.

20. Kumaraswamy E., Mahender K., Prasad C.R., Govardhan N., Yadav B.P. Digital watermarking techniques: Comparative analysis and robustness for real time applications 2022 AIP Conference Proceedings 2418 30070 10.1063/5.0081731.

21. Mohammed Ali Shaik, Time Series Forecasting using Vector quantization, International Journal of Advanced Science and Technology (IJAST), 29(4), 169-175, (2020).

22. P. Praveen, B. Rama and T. Sampath Kumar, "An efficient clustering algorithm of minimum Spanning Tree," in Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB, Chennai, 2017), Pp. 131-135.

23. Kumaraswamy E., Mahesh Kumar G., Mahender K., Bukkapatnam K., Prasad C.R. Digital Watermarking: State of the Art and Research Challenges in Health Care & Multimedia Applications 2020 IOP Conference Series: Materials Science and Engineering 981 3 32031 10.1088/1757-899X/981/3/032031.

24. T. Sampath Kumar, and B. Manjula, Perusing on Cloud Computing and its Security Issues, International Journal of Engineering and Advanced Technology (IJEAT), 9(2), 123-129, (2019).

25. Kumaraswamy E., Sharma S., Kumar S. A review on cancer detection strategies with help of biomedical images using machine learning techniques 2022 AIP Conference Proceedings 2418 30064 10.1063/5.0081936.

26. Prashanth B., Krishna D.B., Balasundaram A., Tejaswi B., Govardhan N. Implementation patterns of high performance machine learning algorithms using Apache Mahout 2022 AIP Conference Proceedings 2418 20044 10.1063/5.0082516.

27. Praveen P., Shaik M.A., Kumar T.S., Choudhury T. (2021) Smart Farming: Securing Farmers Using Block Chain Technology and IOT. In: Choudhury T., Khanna A., Toe T.T., Khurana M., GiaNhu N. (eds) Blockchain Applications in IoT Ecosystem. EAI/Springer Innovations in Communication and Computing. Springer.

28. T.Sampath Kumar, B.Manjula, Asymmetric AES Algorithm for Cloud Security, International Journal of Future Generation Communication and Networking 12(5), 301- 305, (2019).

29. R. Ravi Kumar, M. Babu Reddy and P. Praveen, "A review of feature subset selection on unsupervised learning," in Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB, Chennai, 2017), pp.163-167.

30. Mendu, M., Krishna, B., Sandeep, C.H., Mahesh, G., Pallavi, J. Development of real time data analytics based web applications using NoSQL databases 2022 AIP Conference Proceedings 2418 020038, doi.org/10.1063/5.0082146.

31. Mohammed Ali Shaik, Geetha Manoharan, B Prashanth, NuneAkhil, Anumandla Akash and Thudi Raja Shekhar Reddy, "Prediction of Crop Yield using Machine Learning", in International Conference on Research in Sciences, Engineering & Technology, (AIP Conf. Proc. 2418, 2022), Pp.020072-1–020072-8, doi.org/10.1063/5.0081726.

32. S.Vijayarani, S.Nithya. An Efficient Clustering Algorithm for Outlier Detection, International Journal of Computer Applications. 22-27, (2011).

33. Prashanth B., Krishna D.B., Shaik M.A., Tejaswi B., Kiran K.R. Optimization factors with high performance computing and data science based implementations with metaheuristics 2022 AIP Conference Proceedings 2418 20043 10.1063/5.0082513.

34. XV. Zhang, D. Chao, and J. Wang, Composite Quantization for Approximate Nearest Neighbor Search, in ICML, 1234-1239, (2014).

35. Prashanth B., Neelima G., Dule C.S., Chandra Prakash T., Tarun Reddy S. Data Science and Machine Learning Integrated Implementation Patterns for Cavernous Knowledge Discovery from COVID-19 Data 2020 IOP Conference Series: Materials Science and Engineering 981 2 22004 10.1088/1757-899X/981/2/022004.

36. Y. Chen, T. Guan, and C. Wang, Approximate nearest neighbor search by residual vector quantization, Sensors, 10(12), 11259– 11273, (2010).

37. M. A. Shaik, S. k. Koppula, M. Rafiuddin and B. S. Preethi, (2022), "COVID-19 Detector Using Deep Learning", in International Conference on Applied Artificial Intelligence and Computing (ICAAIC, 2022), pp. 443-449.

38. Sudarshan E., Kumari D.A., Reddy Y.C.A.P., Balasundaram A., Mahender K. Machine learning based automatic vehicle alert system 2022 AIP Conference Proceedings 2418 20058 10.1063/5.0081741.

39. Yu-Chen Song, J.O'Grady, G.M.P.O'Hare, Wei Wang, "A Clustering Algorithm incorporating Density and Direction", IEEE Computer Society, (CIMCA 2008).

40. Mohammed Ali Shaik, Praveen Pappula, T Sampath Kumar, Predicting Hypothyroid Disease using Ensemble Models through Machine Learning Approach", European Journal of Molecular & Clinical Medicine, 9 (7), 6738-6745, (2022).

41. Sudarshan E., Naik K.S., Kumar P.P. Parallel approach for backward coding of wavelet trees with CUDA 2020 ARPN Journal of Engineering and Applied Sciences 15 9 1094 1100.

42. Yadav B.P., Ghate S., Harshavardhan A., Jhansi G., Kumar K.S., Sudarshan E. Text categorization Performance examination Using Machine Learning Algorithms 2020 IOP Conference Series: Materials Science and Engineering 981 2 22044 10.1088/1757-899X/981/2/022044.

43. Yadav B.P., Sheshikala M., Swathi N., Chythanya K.R., Sudarshan E. Women Wellbeing Assessment in Indian Metropolises Using Machine Learning models 2020 IOP Conference Series: Materials Science and Engineering 981 2 22042 10.1088/1757-899X/981/2/022042.

44. Bhushan T., Chandrashekhar A., Venkat Prasat S., Rajasri Reddy I. Effect of Substrate Surface Roughness on Adhesion of Titanium Nitride Coatings Deposited by Physical Vapour Deposition Technique 2020 IOP Conference Series: Materials Science and Engineering 981 4 42022 10.1088/1757-899X/981/4/042022.

45. Rajasri Reddy I., Reddy C.V.K., Rao Y.V.D., Chandra Shekar A. Comparison of Tests for Isomorphism in Planetary Gear Trains 2020 IOP Conference Series: Materials Science and Engineering 981 4 42023 10.1088/1757-899X/981/4/042023.

46. Sammaiah P., Chaitanya Krishna D., Sai Mounika S., Rajasri Reddy I., Karthik T. Effect of the Support Structure on Flexural Properties of Fabricated Part at Different Parameters in the Fdm Process 2020 IOP Conference Series: Materials Science and Engineering 981 4 42030 10.1088/1757-899X/981/4/042030.