



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 8, Issue 12, December 2020

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.488

 9940 572 462

 6381 907 438

 ijirccce@gmail.com

 www.ijirccce.com

A Survey on Prediction of Building cost using k means++ Algorithm in Python Framework

Yogeshwari B¹, Dr.V.Sharmila², Dr.M.Somu³, Dr.V.Vennila⁴, Dr.J.Preetha⁵

PG Student, Dept. of C.S.E, K.S.R College of Engineering (Autonomous), Tiruchengode, Tamilnadu, India¹

Professor, Department of Computer Science & Engineering, K. S. R. College of Engineering (Autonomous),
Tiruchengode, Tamilnadu, India^{2,3,4}

Professor, Department of Computer Science & Engineering, Muthayammal Engineering College, Rasipuram,
Tamilnadu, India⁵

ABSTRACT: This paper is to predict the building cost of various sectors. This Prediction gives more accurate result by using k means++ algorithm. It clearly defines the basic machine learning concept. This algorithm works under clustering concept which is nothing but grouping or segmentation of the data. This algorithm which comes under unsupervised learning gives more accuracy. The proposed factors are used to predict the building cost with analyzing the dataset which we taken from different sources. The most important outcome of this study is to find out the machine learning algorithms to predict the cost estimation of construction. Either overestimating or underestimating the cost of these projects will lead to future deviations in budget vs. realized cost. Hence, the methods used in this realm, their respective accuracy, and even their gaps have shown growing interest.

KEYWORDS: Machine Learning, Clustering, Unsupervised learning, K – means++

I. INTRODUCTION

Adequate construction cost estimation is main factor in any type of construction projects. Forecasting cost of construction projects can be considered as difficult task. Construction cost estimating is the process of forecasting the cost of building a physical structure. Machine learning techniques need adequate dataset size to model and forecast the cost of projects. Cost prediction is a vital process for every business in that it is a predecessor for budget prices and resource allocation in a project life cycle. Actually, it is hard to obtain input data for cost estimation process, while the scope of work is barely known in that it might lead to poor and rough estimates. The more, the project scope is known there are more chances to generate estimates that are more accurate in that more specifications of the project are defined.

II. LITERATURE REVIEW

Ayush Varma[1], Abhijit Sarma[1], Sagar Doshi[1], Rohini Nair[1] stated in their paper as Predicting housing prices with real factors is the main crux of our research project. Here we aim to make our evaluations based on every basic parameter that is considered while determining the price. We use various regression techniques in this pathway, and our results are not sole determination of one technique rather it is the weighted mean of various techniques to give most accurate results. The results proved that this approach yields minimum error and maximum accuracy than individual algorithms applied. P. Durganjali[2], M. Vani Pujitha[2] stated the resale price prediction of the house is done using different classification algorithms like Logistic regression, Decision tree, Naive Bayes and Random forest is used and we use AdaBoost algorithm for boosting up the weak learners to strong learners.

H.Raga Madhuri[3], Anuradha G[3] focused on the selling cost of a house perfectly and to help people to predict the exact time slap to accumulate a house. Some of the related factors that impact the cost were also taken into considerations such as physical conditions, concept and location etc.

Zhongyun[4], Jiang, Guoxin, Shen[4] predict the price of second-hand housing in Shanghai. Firstly, this paper use the crawler technology to parse the URL text information through the json request address and the Beautiful Soup parser. Then a multi-layer feed forward neural network model trained by error inverse propagation algorithm is established based on the deep learning library Keras. Finally, to enter standardized data to predict the price.

YingYu, HuangbaoSong, Tianle Zhou, Hanakiachi, Shangce Gao[5] use DNM to fit the House Price Index (HPI) data and then forecast the trends of Chinese housing market. To verify the effectiveness of the DNM, we use a

traditional statistical model (i.e., the exponential smoothing (ES) model) to make a performance comparison. Ruth Erna Febrita, Adyan Nur Alfiyatin, Hilman Taufiq, Wayan Firdaus Mahmudy[6] extract fuzzy rules, which can be used to predict house prices based on nearby objects location. K-Means clustering method is used to extract initial values to form fuzzy membership functions and inference rules of several groups of residential. This research produces a good-interpretability fuzzy system shows a satisfactory result of predictions. Feng Wang, Yang Zou, Haoyu Zhang and Haodong[7] show that individual house price predicted by the proposed approach is better than that of SVR method. And the predicted house price trend is mainly agreement with the real situation. Then house price trend is predicted based on the ARIMA model.

Suraya Masrom, Thuraiya Mohd, Nur Syafiqah Jamil[8] describes the way to reduce the complicated design is by using Automated Machine Learning. A real dataset of house prices in the area of Petaling Jaya has been conducted to test the performances of AML.

Zhen Peng, Qiang Huang, Yincheng Han[9] study the housing price of second-hand houses, this paper analyzed and studied 35417 pieces of data captured by Chengdu HOME LINK network. Firstly, the captured data were cleaned and the characteristics were selected and also prevents overfitting phenomenon, laying a solid foundation for the subsequent second-hand house price prediction

Rushab Sawant, Yashwant Jangid, Tushar Tiwari, Saurabh Jain, Ankita Gupta[10] has predicted to grow at 30-35% over the next decade. In terms of employment provided, it is second only to the agricultural sector. Housing is one of the major domain of real estate. The interests of both buyer and seller should be satisfied so that they do not overestimate or underestimate price.

V. Sharmila, G. Tholkappia Arasu, P. Balamurugan[11] has proposed a non-class element based iterative clustering approach. In this approach based on the weight calculation classes are selected.

V.Vennila, A. Rajiv Kannan[12] introduced parallel linguistic fuzzy rule with canopy MapReduce (LFR-CM) framework. The framework classifies big data using canopy MapReduce function for information sharing in cloud with higher classification accuracy and lesser time consumption.

P. Balamurugan, T.Ravichandran, V.Sharmila [13] proposed Grade- Based Data Gathering (GBDG) algorithm to minimize the energy consumption of wireless sensor networks.

V.Vennila, A. Rajiv Kannan[14] proposed Discretized Support Vector Classification and Prediction (DSV-CP) model to provide efficient Big Data computation and information sharing in Cloud computing environment.

V.Sharmila, P.Balamurugan, V.Vennila, S.Savitha [15] has proposed a data verification scheme to identify the malicious data packets.

P. Balamurugan, M. Shyamala Devi, V. Sharmila [16] has introduced the optimized methods for securing data (OMSD) which is trust based weights and also completely about the attacks and some methods for secured data transmission

V.Vennila, A. Rajiv Kannan[17] has proposed Parallel Symmetric Matrix-based Predictive Bayes Classifier (PSM-PBC) model is developed for efficient Big Data computation and information sharing in Cloud environment.

P. Balamurugan, M. Shyamala Devi, V. Sharmila [18] has proposed Score-based data gathering algorithm provides a significant solution to maximize the network lifetime as well as minimum delay per round of data gathering.

III. METHODOLOGY

A. Existing System

Stacking algorithm is applied in the existing system. It is difficult task to predict the accurate values of house pricing. The Existing system focuses only for the house cost. It takes more time. The Performance ratio is low when compared to proposed system.

B. Proposed System

In proposed system we are using K means++ algorithm. It focuses on any type of building. It predicts the cost of building based on numerical and geometrical values. In the proposed system we can process more number of data within a short period of time. The performance ratio is very high and it gives more accurate result. Its cost is cheaper. Its execution time is very less. The performance ratio is very high.

C. Data Pre-Processing

Data Collection: Collecting the relevant dataset from UCI data repository. Data are analyzed and outliers are detected and it should be removed to get more accurate result.

IV. CONCLUSION

The prediction report is proposed to effectively predict the data among the dataset. A precise prediction of the housing price is essential to the perspective of house- owners as everything belongs to a real estate market. We test for the performance of these techniques by computing how precisely a technique can predict whether the closing price is

greater than or less than the listing price. The system will satisfy customers by providing accurate output and preventing the risk of investing in the wrong house. This project efficiently analyzes past industry trends and price ranges to predict future prices. To achieve the results, various data mining techniques are utilized in python language. Various factors which affect the building pricing are considered and further worked upon them. Machine learning has been considered to complete out the desired task. Firstly, data collection is performed. Then data cleaning is performed to remove all the errors from the data and make it clean. Then data pre-processing is done. Then with the help of data visualization, different plots are created, which intends to depict the distribution of data in different forms. Towards the end, the business costs of the buildings were determined with exactness and accuracy.

REFERENCES

1. Ayush Varma, Abhijit Sarma, Sagar Doshi, Rohini Nair House Price Prediction Using Machine Learning And Neural Networks, IEEE – 1st Sep, 2020.
2. P. Durganjali, M. Vani Pujitha House Resale Price Prediction Using Classification, ICSSS, 27th Feb, 2019.
3. H.Raga Madhuri, Anuradha G House Price Prediction Using Regression Techniques: A Comparative Study, ICSSS, 2019.
4. Zhongyun, Jiang, Guoxin, Shen Prediction of House Price Based on The Back Propagation Neural Network in The Keras Deep Learning Framework, ICSAI, 2019.
5. YingYu, HuangbaoSong, Tianle Zhou, Hanakiachi, Shangce Gao Forecasting House Price Index of China Using Dendritic Neuron Model, IEEE, 2016.
6. Ruth Erna Febrita, Adyan Nur Alfiyatin, Hilman Taufiq, Wayan Firdaus Mahmudy Data-driven Fuzzy Rule Extraction for Housing Price Prediction in Malang, East Java, ICACSSIS, 2017.
7. Feng Wang, Yang Zou, Haoyu Zhang and Haodong House Price Prediction Approach based on Deep Learning and ARIMA Model, ICCSNT, 2019.
8. Suraya Masrom, Thuraiya Mohd, Nur Syafiqah Jamil Automated Machine Learning based on Genetic Programming: a case study on a real house pricing dataset, IEEE, 2019.
9. Zhen Peng, Qiang Huang, Yincheng Han Model Research on Forecast of Second-Hand House Price in Chengdu Based on XGboost Algorithm, ICAIT, 2019.
10. Rushab Sawant, Yashwant Jangid, Tushar Tiwari, Saurabh Jain, Ankita Gupta Comprehensive Analysis of Housing Price Prediction in Pune using Multi-Featured Random Forest Approach, IEEE, 2018.
11. V. Sharmila, G. Tholkappia Arasu, P. Balamurugan Non-Class Element based Iterative Text Clustering Algorithm for Improved Clustering Accuracy using Semantic Ontology, Asian Journal of Research in Social Sciences and Humanities, 2016.
12. V.Vennila & Dr. A. Rajiv Kannan Hybrid Parallel Linguistic Fuzzy Rules with Canopy MapReduce for Big Data Classification in Cloud, International Journal of Fuzzy Systems(IJFS), January 2019.
13. P. Balamurugan, T.Ravichandran, V.Sharmila Grade and Energy based Data Gathering Protocols in Wireless Sensor Networks, Asian Journal of Research in Social Sciences and Humanities, 2016.
14. V.Vennila & Dr. A. Rajiv Kannan Discretized Support Vector Prediction Classifier for Big Data Computation and Information Sharing in Cloud, Asian Journal of Research in Social Sciences and Humanities, 2017.
15. V.Sharmila, P.Balamurugan, V.Vennila, S.Savitha Information Retrieval and Recommendation Framework Using Maximum Matched Pattern Based Topic Models, International Journal of Innovative Research in Engineering Science and Technology, 2016.
16. P. Balamurugan, M. Shyamala Devi, V. Sharmila Detecting malicious nodes using data aggregation protocols in wireless sensor networks, International Journal of Engineering & Technology, 2018.
17. V.Vennila & Dr. A. Rajiv Kannan Symmetric Matrix-based Predictive Classifier for Big Data computation and information sharing in Cloud, Elsevier Computers and Electrical Engineering, 2016.
18. Balamurugan, M. Shyamala Devi, V. Sharmila An energy minimizing score based optimal data gathering in wireless sensor networks, International Journal of Engineering & Technology, 2018.
19. Somu, M. and Rengarajan, N. 'An Improved particle swarm Optimization Based on Deluge Approach for Enhanced Hierarchical cache Optimization in IPTV Networks' Research Journal of Applied Sciences, Engineering and Technology, 2014.
20. Somu, M. and Rengarajan, N. 'A Hybrid Model of Swarm Intelligence Algorithm to Improve the Hierarchical Cache Optimization in IPTV Networks' International Review on Computers and Software, 2013.
21. Somu, M. and Rengarajan, N. (2013) 'A Replacement Policy for Buffer Management in IPTV Services' ISSN 2320-088X IJCSMC- International Journal of Computer Science and Mobile Computing, Vol. 2, Issue. 6, June 2013, pg.140 – 144.
22. Somu, M. and Rengarajan, N. (2012) 'A Review on the Performance of Caching Algorithms for Video Streaming Services in IPTV' International Journal of Advanced Research in Computer Science and Software Engineering ISSN: 2277 128X Volume 2, Issue 5, pp.350-355.
23. Somu, M and Dineshkumar K (2015), 'A survey on distribution database management for minimizing the energy consumption in wireless sensor' published in international journal of modern trends in engineering and research(IJMT ER) Journal Vol. 2, Issues 9. Pp.112-116



INNO  SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details