JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

PREDICTION OF REAL ESTATE PRICE USING DATA MINING TECHNIQUES

Charow N
PG Scholar, Master of Computer Applications Hindustan Institute of Technology
Chennai, India Charow939@gmail.com

Dr. Sudha
Professor,Dept. of Computer Applications Hindustan Institute of Technology Chennai, India Sudhas@hindustanuniv.ac.in

***Abstract*—Data Mining plays a vital role in prediction model development and knowledge discovery. This project follows the Cross-Industry Standard Process for data processing so as to develop a prediction model for the real estate land prices in numerous cities and villages within the Tamilnadu. the info has been collected from different sources which include social media networks, official newspapers, and advertisement newspapers, then it's been analyzed using WEKA data mining tool for model development and analysis. the target of this project is to use data mining to assist the real estate advisors and consumers within the pricing estimations of the real estate land prices, and to cope with any changes within the and costs consistent with the market trends. The project considers variety of things which affect the land price, such as: land length and width, location, and land classification. linear regression is applied for analysis. The study shows that the model is influenced by the variance of the land price between different areas within the Tamilnadu, and so variety of enhancements are suggested for future development.**

I. INTRODUCTION

The powerful prediction models developed by Data Min- ing (DM) tools and algorithms increased the confidence of businesses in the application of DM. Prediction models are currently used to solve forecasting issues, consumer behaviour, patterns prediction, and market analysis. This Project looks into the implementation of DM in the real estate field, in order to develop a prediction model which is able to estimate the land prices in different areas within the Tamilnadu. The dataset used in this research has been collected from different sources, and contains 120 real estate transactions. The Project follows a standard DM process, which is the Cross-Industry Standard Process for Data Mining (CRISP-DM) to analyse the collected data, in order to generate a beneficial prediction model for the real estate advisors and consumers to help them having a better understanding of the market trends and an estimated market price of the targeted land of interest. A solid prediction model with reliable accuracy of prediction is the main target of this study. The first section of paper will demonstrate the research methodology and approach followed, which will include a background of real estate in Tamil Nādu and the current market status. This section will discuss the DM process including the dataset description, pre-processing steps, and the use of Linear Regression as a DM method with the results that have been obtained. The study concludes with a number of suggestions for a better model prediction.

II. LITERATURE SURVEY

1. Data processing techniques are broadly classified intotwo classes Statistical Techniques and (ii) Knowledge Discovery. The continuing rapid growth of on-line data and also the widespread use of databases necessitate the development of techniques for extracting useful knowledge and for facilitating database access. This paper Analysis of data Mining Tech- niques on real estate analyzes the results of multilayer percep- tron with pace regression and suggests a very efficient pattern which might be proved beneficial for knowledge discovery. The analysis is finished using property data set which contains 5821 tuples and 43 attributes and determines that in India's scenario the demographic details of someone plays a very prominent role in identifying the investment behaviour of a customer. If we are discarding the demographic details then the model which is available consists of 13 Sigmoid nodes and there is a significant change in error rate and correlation. we've used WEKA for analysis and found that generally multilayer perceptron(selected) is more efficient than pace regression(complete) in terms of statistical methods, but in Indian perception pace regression(complete) is more efficient.

2. This Paper Empirical Application in Real Estate Valua- tion presents the insights gained from applying data mining techniques, in particular neural networks for the purposes of developing an intelligent model used to predict real estate property values based on variety of factors. A dataset of over one thousand transactions in real estate properties was used. The dataset included 15 variables obtained from the multiple listing system (MLS) database and captured information on transactions taking place during a period of three years. The results from applying data mining techniques to predict real

estate values are promising. Future plans and recommenda- tions for further expanding the study are given.

3. This Paper A Data Mining Model by Using ANN for Predicting Real Estate Market aims to demonstrate the importance and possible value of housing predictive power which provides inde- pendent real estate market forecasts on home prices by using data mining tasks. A (FFBP) network model and (CFBP) network model are one of these tasks used in this research to compare results of them. We estimate the median value of owner occupied homes in Boston suburbs given 13 neighborhood attributes. An estimator can be found by fitting the inputs and targets. This data set has 506 samples. "Housing inputs" is a 13 × 506 matrix. The "housing targets" is a 1 ×506 matrix of median values of owner-occupied home and the percent correctly predicting the simulation sample is 96

4. This Paper Finding Interest of People in Purchasing Real Estate using Data Mining Techniques presents that Data mining is the extraction of hidden predictive information from large databases; it is a powerful technology with great potential to help organizations focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, helps organizations to make proactive knowledge-driven decisions. Hence by using data mining techniques we predict the interest of people in real estate and their pattern of purchasing them. The data has collected by moving the questionnaire among the people. We used two data mining techniques that classify the data based on certain attributes, are classification (Zeror classifier) and clustering (simple k means).

III. Existing System

The Existing System analysed the demographic details for the behaviour of the customers' investments using Artificial Neural Networks (ANN) and regression. This research study was more focused on evaluating the effectiveness of each DM technique in handling the attributes used in the dataset which is not in lined with our objective of the research. The study used a dataset of around a thousand of real estate transactions that have been collected over three years in the state of Tamil Nadu. Jean's research study focused on the prediction of the real estate properties, and did not include a prediction for land values. The research study indicates a successful prediction model using ANN; however, it does not include any results analysis. In addition to that, the study focused on different types of property rather than land prices which is the objective we are aiming for.

Disadvantages:

i. Less Accuracy ii. Variance is high iii. Primarily not focused on Land price analysis

IV. Proposed System

This proposed system of this project is based on prediction of land prices that is generated by Data Mining Technique. We utilize Linear regression as our model because of its adaptable and probabilistic methodology on model selection. Our result exhibit that our approach of the issue need to be successful, and has the ability to process predictions that would be comparative with other Land price prediction models. This has imposed the problem of overpricing or low pricing for some of the real estate lands, with no solid estimation in place. This Project uses DM in order to solve the reduction problem of the real estate land price.

Advantages :

i. High Accuracy ii. Ease to use iii. Minimal Variance

Methodology and Approach

A. Background of the Research Study

The boost in Tamil Nādu economy has made the real estate values of land prices booming a year after year. In 2010, Tamil Nādu Survey and Land Registration Bureau (SLRB) stated that the land prices have been increased more than 300

B. Dataset Description

There are some factors and variables which affect the price estimation of land prices in Tamil Nādu. There big differences among land prices located in Chennai and other places like Tindivanam. The classification category by the SLRB and the number of roads also play a role in the land price.

C. Data Pre-processing

Since there is a large number of cities and villages in Tamil Nādu, the research study focuses on 10 selected locations only for the prediction model. There are many land classification categories such as: RHA, RA, RB, B3, B4. . . etc. In this study, we are only interested to predict the RA and RB categories. Instances with outliers have been removed. The dataset includes 120 instances with no missing values or outliers.

D. Prediction Algorithm

There are many prediction algorithms which can be used for prediction of numerical values in a dataset such as ANN and LR, which are the most well-known algorithms. In this study we have implemented LR for the development of the prediction model.

E. DM Analysis and Results

Based on the analysis produced by WEKA, it shows clearly that the LR model analysis on the dataset is not promising. No solid regression has been discovered on the dataset. The model shows a relative absolute error of more than 77

V. Modules

A. Module1 : User 1

The user1 has Create a new account.The user1 has Login with their credential details.The user1 can register their Land details to the database.

B. Module2 : User 2

The admin has Login with their credential details.Admin can view the user details and land details from the database.Admin can predict the land price as per the user1 location.

C. Module3 : User 3

The user2 has create a new account.The user2 has Login with their credential details.User2 can view the land details from the database if they need to buy the land.

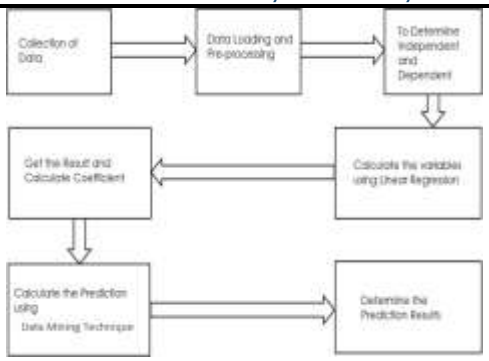

Fig. 1.

VI. DATA FLOW DIAGRAM

VII. CONCLUSION

Further work can be achieved regarding the research study on the real estate prediction model, a number of enhancements and input parameters can be added to make the model more robust. For example, as the land location is not an accurate identifier, we suggest to add the (x,y) coordinates of the land. The date of the real estate transaction is an important factor as well, prices are under market change from time to time, hence having the transaction date will help. The model needs to have a real time data, to cope with the market change and demand for a better prediction analysis. So, it is recommended to keep the dataset up to date, otherwise it does not reflect the actual market price.

REFERENCES

[1]G. Banerji and K. Saxena, Analysis of Data Mining Techniques on Real Estate, International Journal of Soft Computing and Engineering (IJSCE), 2012, Vol. 2.

[2]Jaen, Data Mining: An Empirical Application in Real Estate Valuation, Association for the Advancement of Artificial Intelligence, 2002, Vol. 2.

[3]I. Bahia, A Data Mining Model by Using ANN for Predicting Real Estate Market: Comparative Study, International Journal of Intelligence Science, 2013, Vol. 3, pp. 162-169.

[4]I. Bahia, A Data Mining Model by Using ANN for Predicting Real Estate Market: Comparative Study, International Journal of Intelligence Science, 2013, Vol. 3, pp. 162-169.

[5]S. Jamil, T. Mohd, S. Masrom and N. Ab Rahim, Machine Learning Price Prediction on Green Building Prices, 2020 IEEE Symposium on Industrial Electronics Applications (ISIEA), Malaysia, 2020, pp. 1-6.

[6]Z. Jiang and G. Shen, Prediction of House Price Based on The Back Propagation Neural Network in The Keras Deep Learning Framework," 2019 6th International Conference on Systems and Informatics (ICSAI), Shanghai, China, 2019, pp. 1408-1412.

[7]F. Wang, Y. Zou, H. Zhang and H. Shi, House Price Prediction Approach based on Deep Learning and ARIMA Model, 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 2019, pp. 303- 307.

[8]W. Qin, A Study on Real Estate Price by Using Probability Statis- tics Theory and Grey Theory, 2019 18th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), Wuhan, China, 2019, pp. 153-156.

[9]C. R. Madhuri, G. Anuradha and M. V. Pujitha, House Price Prediction Using Regression Techniques: A Comparative Study, 2019 International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2019, pp. 1-5.

[10]State of the real estate market, Alayam Newspaper, accessed on: 27- May-2014, http://www.alayam.com/News/economic/Economic/182410.