International Journal of Computer & Mathematical Sciences
IJCMS
ISSN 2347 – 8527
Volume 6, Issue 12
December 2017

A Multi Feature Based Housing Price Prediction for Indian Market Using Machine Learning

Rushab Sawant[1], Saurabh Jain[2] , Tushar Tiwari[3], Yashwant Jangid[4], Ms.Ankita Gupta[5]

[1,2,3,4,5]AISSMS College of Engineering

ABSTRACT:

The housing sector is the second largest employment provider after agriculture sector in India and is estimated to grow at 30% over the next decade. Housing is one of the major sectors of real estate and is well complemented by the growth of the urban and semi-urban accommodations. Pune is emerging as one of the major metropolitan cities of India. Ambiguity among the prices of houses makes it difficult for the buyer to select their dream house. The interest of both buyers and sellers should be satisfied so that they do not overestimate or underestimate price. Our system provides a decisive housing price prediction model to benefit a buyer and seller or a real estate agent to make a better-informed decision system on multiple features. To achieve this, various features are selected as input from feature set and various algorithms can be applied such as Random Forest and SVM.

Keywords: Random Forest, Supervised Learning, Feature Extraction, Ensemble learning.

1 INTRODUCTION

1.1 BACKGROUND

Food, clothing, and housing (shelter) are the primary requirements of life. The availability of these requirements increases the physical efficiency and productivity of the people. So housing is a factor of prime importance in human resource development. At one point in life, everybody has to deal with the housing dilemma. For many people housing is one of the major investment of their life, people pay their fortune to buy the Dreamhouse.

1.2 About Pune - the city under current analysis :

Pune is the second-largest city in Maharashtra, India and the ninth largest city in the country. Pune has a population density of 5,600 people per square kilometer (15,000/square mile).[15] With a population of over 5 million, Pune is the 9th largest metropolitan area in India. It is one of the fastest-growing cities in the Asian-Pacific. Between 1991 and 2001, the city grew by 40% increase from 1.6 million to 2.5 million.[15] The decadal growth rate of Pune for the last 40 years has been at least 40%. The city has the nickname as "Oxford of the East" because of various Universities and students coming from all parts of India. Pune has many job opportunities, courtesy of its ever-growing IT Parks opening up in special economic zones such as Hinjewadi.

Pune is included among 20 smart cities, boosting the city's image and this has gained a positive impact on the real estate market in both residential and commercial sectors. The high demand for homes by IT professionals, BFSI and manufacturing companies, has launched many residential projects in Pune. The growing housing demand in Pune is attracting large investments in the city and many new projects are being launched.[15]

1.3 Problem under consideration :

Housing for All by 2022 project by Union government pushes the investment towards an affordable housing. Government policies like the Real Estate Regulatory Authority (RERA) have influenced fresh buyer condense into the real estate sector. First-time home-buyers are finally making purchase decisions for homes that meet their budgets. The availability of cheap Nance is also driving the demand for a portable housing. Pune real estate unsold inventory has reached 2.8 lakhs It might seem strange that massive demand and massive unsold residential inventory can co-exist, but this is the result of pricing mismatch.[14]

Day by day the gap between massive demand and massive unsold residential inventory is increasing considerably. Price of the house can't be known to a buyer since it renders taking into account many features

which is too complex and time-consuming. This works as an advantage to the agents and they manipulate the price and increase their own profits. The home buyers are exploited, as their investments are ruined. Thus creating a fear in the peer of buyers, that buying a house is a risky decision. Then, in turn, the houses of the developers are unsold in the market. Hence a great market is turned into a bad market because of the inappropriate pricing of the houses and the greed of the middleman that is agents.

1.4 Addressing the Problem :

In our current project, we are developing a better, efficient and accurate house price prediction system. It is a supervised learning method in which we are following a sequence of steps: First step is Data Collection, the second step is Developing a model for the dataset, the third step is finding out the relationship among the various features and its respective price. And last is to predict the price of a particular house. In this study, we will try to find out more accurate house prices based on various features that influence the Indian buyer's decision. We will try to predict the area to the developers which will maximize their investment returns for their future projects.

1.5 Solution that we aim to formulate :

The aim of the project is to predict the prices of houses in Pune City, to ease and facilitate the identification of housing prices so that both buyer and seller are satisfied.

2 RELATED WORK

2.1 Nissan Pow, Emil Janulewicz, Liu (Dave) Liu, 2016, "Applied Machine Learning Project 4 Prediction of real estate property prices in Montreal "

Applied Machine Learning Project 4 Prediction of real estate property prices analyzed the real estate property prices in Montreal. The information on the real estate listings was extracted from Centris.ca and duProprio.com. They predicted both asking and sold prices of real estate properties based on features such as geographical location, living area, and number of rooms, etc. Additional geographical features such as the nearest police station and fire station were extracted from the Montreal Open Data Portal, the final price sold was also predicted with an error of 0.023 using the Random Forest Regression.[4]

2.2 Itedal Sabri Hashim Bahia, 2013, "A Data Mining Model by Using ANN for Predicting Real Estate Market: Comparative Study"

A (FFBP) network model and (CFBP) network model are one of these tasks used in Data Mining Model by Using ANN to compare results of them. Itedal Sabri Hashim Bahia concludes that which one of the two networks appears to be a better indicator of the output data to target data network structure than maximizing predict.Paper aims to demonstrate the importance and possible value of housing predictive power which provides independent real estate market forecasts on home prices by using data mining tasks.[11]

2.3 Yu, Jiafu Wu. 2016, "Real Estate Price Prediction with Regression and Classification"

Yu, Jiafu Wu. has predicted house prices given explanatory variables that cover many aspects of residential houses. As continuous house prices, they are predicted with various regression techniques including Lasso, Ridge, SVM regression, and Random Forest regression; as individual price ranges, they are predicted with classification methods including Naive Bayes, logistic regression, SVM classification, and Random Forest classification.[1]

2.4 Da-Ying Li, Wei Xu, Hong Zhao, Rong-Qiu Chen, 2009.A SVR Based Forecasting Approach for Real Estate Price Prediction

Da-Ying Li proposes support vector regression (SVR) to forecast real estate prices in China. Aim of the paper is to examine the feasibility of SVR in real estate price prediction. To achieve the aim, five indicators are selected as the input variables and real estate price is used as output variable of the SVR. The quarterly data during 1998-2008 are employed as the data set to construct the SVR model. With the scenarios, real estate prices in future are forecasted and analyzed. The forecasting performance of SVR model was also compared with BPNN model. [10]

International Journal of Computer & Mathematical Sciences
IJCMS
ISSN 2347 – 8527
Volume 6, Issue 12
December 2017

2.5 L.Li, K.-H. Chu, 2017,Prediction of Real Estate Price Variation Based on Economic Parameters

In Prediction of Real Estate Price Variation Based on Economic Parameters L.Li, K.-H. Chu, has used macroeconomic parameters on real estate price variation are investigated before establishing the price fluctuation prediction model. Here, back propagation neural network (BPNN) and radial basis function neural network (RBF) two schemes are employed to establish the nonlinear model for real estates price variation prediction of Taipei, Taiwan based on leading and simultaneous economic indices. The public related data of Taipei, Taiwan real estate variation during 2005-2015 are adopted for analysis and prediction comparison.[3]

3 RELATED TERMS

3.1 SUPERVISED LEARNING

Supervised machine learning is a machine learning algorithm that uses a labelled dataset for prediction. Labelled Data means an output is associated with every input. In other words, supervised learning builds a model using this dataset that can make a prediction of the new or unseen dataset. This unseen or new dataset is called training dataset and helps in validating the model. Supervised algorithms can be classification algorithms like Support Vector Machines, Naive Bayes, Decision trees or regression algorithms like linear regression, neural networks, and decision trees. Since decision trees can be used for both classification and regression, they are one of the best known supervised algorithms.

3.2 ENSEMBLE LEARNING

It is a machine learning paradigm in which multiple weak learners are trained and combined to form a strong learner. It combines a diverse set of individual learners together to improvise on the predictive power and stability of the model. A number of learners (base learners) form an ensemble. The ensemble has the stronger ability for generalization than that of individual base learners. Ensemble learning is a better choice because it has the ability to boost weak learners which are slightly better than a random guess to strong learners which can make very accurate predictions. “Base learners” are also called “weak learners”.

On training by a base learning algorithm like a decision tree, neural network or others, base learners are generated. Ensemble methods generate base learner in two ways:

1) Homogeneous base learners, in which same base learning algorithm is applied to generate individual learners.

2) Heterogeneous base learners, in which different base learning algorithms are applied to generate individual learners.[13]

3.3 RANDOM FOREST

A combination of many tree predictors (here decision tree) such that each tree is generated independently from another tree. As the number of trees in the forest increases the generalization error for the forest converges.

Factors on which generalization error depends are i) strength of the individual trees and ii) correlation among the trees in the forest.[12]

Definition : A random forest is a classifier consisting of a collection of tree-structured classifiers $\{h(\ , \ _k),k=1…..\}$ where the $\{ \ _k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x .

3.4 FEATURE EXTRACTION

Sometimes, a dataset may include some features that may not be as important as other features. Such features do not help much in classifying data or sometimes may lead to low accuracy. It is important to select the best features so that a better and efficient model is created. Using same features, again and again, can cause overfitting and underfitting of data. Techniques such as k-fold cross-validation can be used to solve overfitting. This will help in constructing a better model. Prediction of the house price is dependent on some of the major features of the house and few of them are:

1)Location.

2)Parking.

International Journal of Computer & Mathematical Sciences
IJCMS
ISSN 2347 – 8527
Volume 6, Issue 12
December 2017

3)Amenities.
4)Stamp Duty Rate.
5)No of rooms.
6)Nearby Places(Hospitals, Railway Station,Gardens)
7)Facing towards.

4 SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

The overall system design consists of following modules:

(a) Data Collection.
(b) Preprocessing
(c) Data Classification.
(d) Data regression.
(e) Prediction of Output.

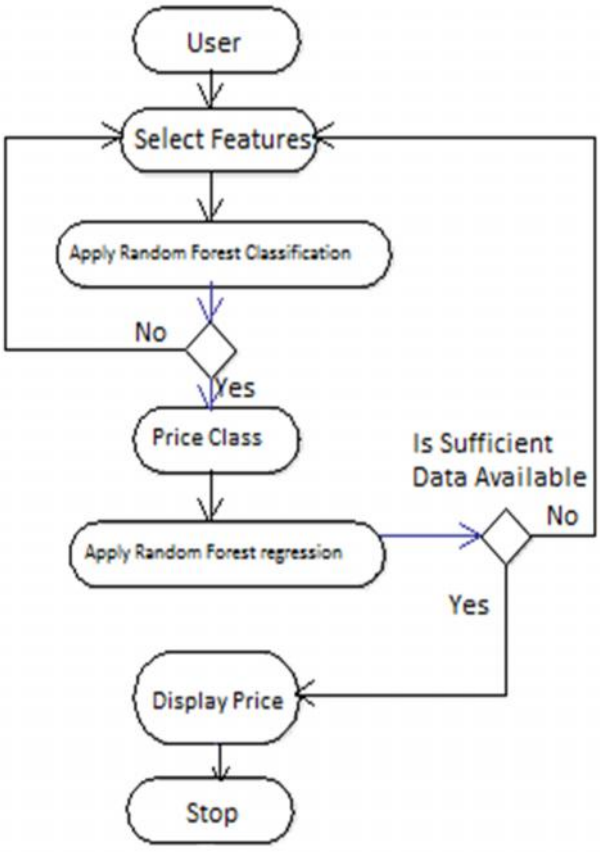

Figure 3.1: System Flow Diagram

Data is collected from various data sources, that data can be structured or unstructured format. For structured data SQL techniques are used to extract data and For unstructured data NOSQL techniques are used to extract data. Classification and then regression algorithms can be applied for price prediction on data. The overall flow of the system is shown in Figure 3.1.

International Journal of Computer & Mathematical Sciences
IJCMS
ISSN 2347 – 8527
Volume 6, Issue 12
December 2017

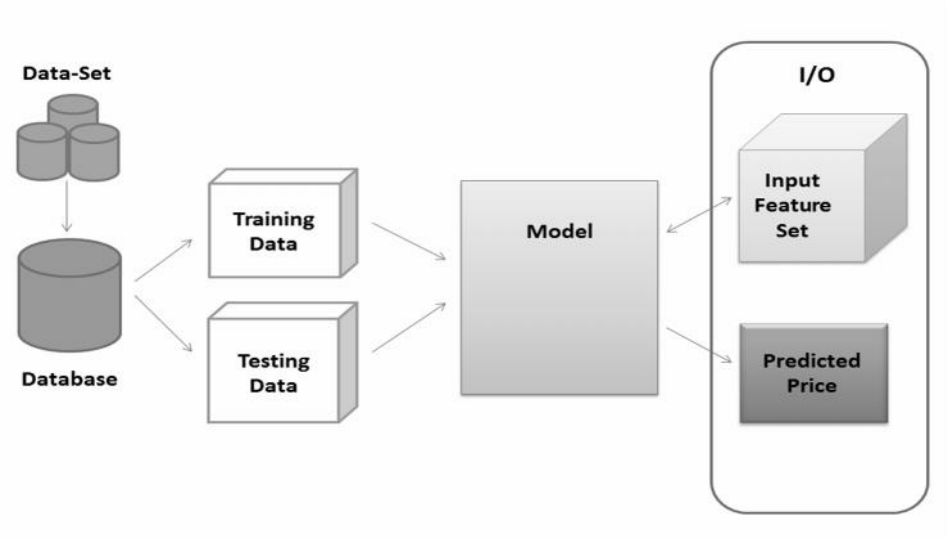

Figure 3.2: Block Diagram

Data is collected and stored in NoSQL / SQL format. That data is divided into two parts

i) Training data

ii) Testing data.

Training data is used for training the model and then that model is tested using testing data. After this, the trained model is used for predicting house price given feature set.

The overall Blocks of the system is shown in Figure 3.2.

5 CONCLUSION

The development of our project till now is just bound to predicting housing prices based on features that do not change with time. In addition to these features, there are various other factors in the market that affect the prices. Parameters like Economy, the inflation rate of an area may result in increase or decrease in the prices. The further project development will be focusing on including these features thereby giving a more precise prediction of prices.

6 REFERENCES

[1] Yu, Jiafu Wu. 2016. Real Estate Price Prediction with Regression and Classification, CS 229 Autumn 2016 Project Final Report, Stanford University.

[2] Y.-C. Wang, R. Huang, C.-C. Nieh, H.-K. Ou, and M. Chi, 2017. Integration between real estate market and the stock market: Evidence from Taiwan, 2017 International Conference on Applied System Innovation (ICASI).

[3] L.Li, K.-H. Chu, 2017. Prediction of real estate price variation based on economic parameters, 2017 International Conference on Applied System Innovation (ICASI).

[4] Nissan Pow, Emil Janulewicz, Liu (Dave) Liu, 2016. Applied Machine Learning Project 4 Prediction of real estate property prices in Montreal: An application of Random Forest, McGill University.

[5] B. Trawinski, Z. Telec, J. Krasnoborski, M. Piwowarczyk, M. Talaga, T. Lasota, and E. Sawilow,2017. Comparison of expert algorithms with machine learning models for real estate appraisal, 2017 IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA).

[6] E. Antipov and E. Pokryshevskaya,2010. Mass Appraisal of Residential Apartments: An Application of Random Forest for Valuation and a CART-Based Approach for Model Diagnostics, SSRN Electronic Journal.

[7] J.-G. Liu, X.-L. Zhang, and W.-P. Wu, 2006. Application of Fuzzy Neural Network for Real Estate Prediction, Advances in Neural Networks - ISNN 2006 Lecture Notes in Computer Science, pp. 1187-1191.

[8] E. Hromada, 2015. Mapping of Real Estate Prices Using Data Mining Techniques, Procedia Engineering, vol. 123, pp. 233-240.

[9] Byeonghwa Parka Jae, Kwon Bae, 2014.Using machine learning algorithms for housing price prediction: The case of Fair-fax County, Virginia housing data: Expert Syst.Appl.42(6),2928-2934(2014)

[10] Da-Ying Li, Wei Xu, Hong Zhao, Rong-Qiu Chen, 2009.A SVR based forecasting approach for real estate price prediction,*International Conference on Machine Learning and Cybernetics*, Baoding.

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

[12] Leo Breiman,2001. RANDOM FORESTS, Statistics Department University of California Berkeley, CA 94720.

[13] Zhou ZH. ,2015. Ensemble Learning. In: Li S.Z., Jain A.K. (eds) Encyclopedia of Biometrics. Springer, Boston, MA.

[14] The Economic Times(21-Sept-2017). Retrieved from https://economictimes.indiatimes.com/news/economy/policy/government-announces-new-ppp-policy-for-private-investments-in-affordable-housing/articleshow/60777583.cms

[15] World Population Review (2017). Retrieved from http://worldpopulationreview.com/world-cities/pune-population/