House Resale Price Prediction Using Classification Algorithms

[1]P. Durganjali, [2]M. Vani Pujitha
[1]M.Tech student, [2]Assistant Professor
Department of Computer Science and Engineering
V R Siddhartha Engineering College, Vijayawada.
anju.panjali520@gmail.com, pujitha@vrsiddhartha.ac.in

Abstract— **Now a days house resale is majorly seen in metro cities. The market demand for housing is always increasing every year due to increase in population and migrating to other cities for their financial purpose. Prediction of house resale price for long-term temporary basis is important especially for the people who stays who will stay the long time period but not permanent and the people who do not want to take any risk during the house construction. In this paper, 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. Several factors that are affecting the house resale price includes the physical attributes, location as well as several economic factors persuading at that time. Here we consider accuracy as the performance metrics for different datasets and these algorithms are applied and compared to discover the most appropriate method that can be used the reference for determining the resale price by the sellers.**

***Index Terms*— House Resale Price, Prediction, Random Forest, AdaBoost, Naïve Bayes.**

I. INTRODUCTION

The value of a home is well known as a combination of a large variety of options. Therefore, the prediction of home price presents a novel set of challenges. Though an oversized variety of space units are dedicated to the present task, their performance and applications are restricted by a really long delay within the handling of data, the dearth of real-world settings and so the inadequacy of choices for housing. Our aim is to predict the value of the house marketing mistreatment classification techniques during this paper. Most of the present studies have concentrated on breakdown the distraction of the prediction of house costs. Several theories are born as an outcome of the analysis work committed by completely different researchers around the world. This paper picks up the most recent prediction analysis to assist economic predictors to use it. It provides a summary of the prediction markets and, together, the present markets that build it easier to predict the market.

II. RELATED WORK

Sifei Lu et.al, [1] introduced a hybrid model for the regression of Lasso and Gradient to predict the price of the individual home. This approach has recently been used as the key kernel for the Kaggle Challenge “House prices: Advanced techniques for regression”.Muhammad Fahmi Mukhlishin et.al, [2] uses several methods to predict the value of land and house. This paper compares Fuzzy logic, Artificial Neural Network, and K-Nearest Neighbor to find the most appropriate method to determine the sellers ' price. Atharva choogle et.al, [3] House price forecasting has been introduced using data mining techniques. It provides a description of the prediction markets and also the current markets that help to make useful predictions in understanding the market. It is therefore, necessary to predict the efficient pricing of real estate customers for their budgets and priorities. Ruth Erna Febrita et.al,[4] presents a Data driven Fuzzy Rule Extraction for Prediction of Housing Prices in Malang, East Ja va In this way, the K Means clustering method is used to extra ct initial values to form fluid membership functions and infere nce rules for several residential groups.The main objective of Lim WT et.al, [5] is to compare the predictive performance o f artificial neural networks with auto regressive integrated mov ing average and multiple regression analysis for condominium prices in Singapore.Tan F, Cheng C et.al, [6] presents a Time aware Latent Hierarchical Model for Predicting House Prices In this context, they introduced a latent hierarchical time aware model to capture the underlying spatiotemporal interactions b ehind the evolution of house prices. Banerjee D et.al, [7] introduced to Predicting the Housing value prediction exploitation machine Learning Techniques. During this, the performance of the machine learning techniques is measured by the four parameters of accuracy, precision, specificity, and sensitivity. The work considers 2 distinct values zero and one as individual categories. Wang JJ et.al, [8] had Predicting House value with a memristor-based Artificial Neural Network. In this, they'd designed a man-made neural network supported memristors is to find out a multivariable regression model with back-propagation formula. Febrita RE et.al, [9] presents Data-driven Fuzzy Rule Extraction for Housing value Prediction in Malang, East Java. In this, they're used KMeans agglomeration technique to extract initial values to make fuzzy membership functions and abstract thought rules of many teams of residential.

978-1-7281-0027-2/19/$31.00 ©2019 IEEE

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on October 04,2024 at 04:11:55 UTC from IEEE Xplore. Restrictions apply.

Yoon JH et.al, [10] introduced Dynamic Demand Response Controller supported time period retail value for residential buildings. During this, they're focussed on developing an effective strategy for the HVAC to reply to time period costs for peak load reduction.

Abbasov C et.al,[11] presents the Prediction of the possibility of commerce of homes because of the issue of monetary stability. During this, they will calculate G score and located that if the computed G price is adequate tozero.39 and bigger than that one it's classified because of the high probability of commerce the house however within the case of being but zero.39 it's classified as none commerce probability.

III. DATA SET

The resale price data of the house was taken from the website of Kaggle. This data set contains various parameters such as house price, storage range, floor size details, etc.

TABLE I
SPECIFICATIONS OF DATA SET

Field name	Value
Number of Rows	217006
Number of columns	13(including Target)
Number of Numeric variables	4
Number of categorical variables	6
Number of Date formats	3
Type of problem	Classification
Number of target variables	2
Missing Values	NIL
Choice of metric	Accuracy

IV. PROPOSED METHODOLOGY

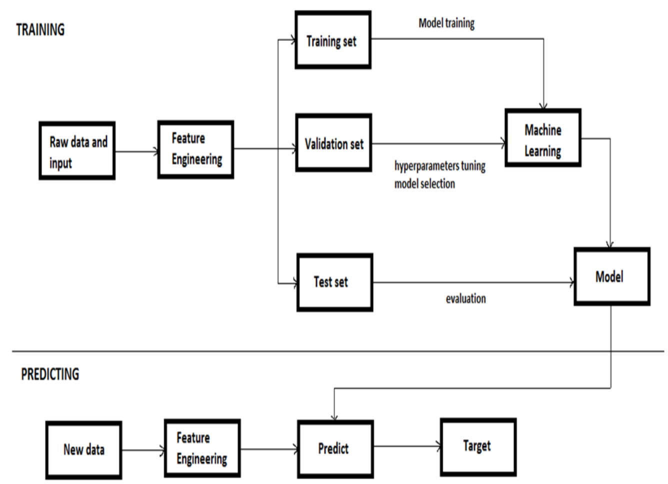


Fig. 1. Structure of Proposed Methodology

The proposed methodology consists of three phases.

- Pre-processing
- Modeling
- Resale price prediction

Phase 1: Pre-processing

In this phase, we carry out the encode variables, outer values, input missing values and take every feasible action that can eliminate disparity in the data set and it partition the given data set into a training data and a test data.

Phase 2: Modeling

This phase uses different classification algorithms like Decision tree, Random forest, AdaBoost, Naïve Bayes, Logistic regression etc. These algorithms give better performance for classification.

Phase 3: Price Prediction

After we get the results from classification we predict the house sale prices and result analysis will also be performed.

Decision Tree Classification

Step1: Decision tree uses multiple algorithms to decide to split a node into two or more sub-nodes.

$$Gain(A) = Info(D) - Info_A(D)$$

Step2: Then splits the nodes further if the node is not classified.

Naive Bayes Classification

978-1-7281-0027-2/19/$31.00 ©2019 IEEE

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on October 04,2024 at 04:11:55 UTC from IEEE Xplore. Restrictions apply.

Step 1: Convert the data set into a frequency table

Step 2: Create a Likelihood table by finding the probabilities

Step 3: Now, use a Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

Ada boost Classification

.

Step 1: It fits a sequence of weak learners on different weighted training data.

Step 2: It starts by predicting the original data set and gives equal weight to each observation.

Step 3: If the prediction is incorrect using the first learner, then it gives higher weight to an observation which has been predicted incorrectly.

Step 4: Being an iterative process, it continues to add learner(s) until a limit is reached in the number of models or accuracy.

Random forest Classification

Step 1: At every node, all of the features are evaluated for the best split.

Step 2: When a new unseen sample is tested, it traverses each of these nodes until it ends in the last node and the average price of all the samples in that node is considered the predicted price.

Step 3: Each node always splits into two other nodes. The depth of tree (levels) can be specified.

Step 4: The minimum number of samples is presented in each node of the tree.

Step 5: max features is another parameter that can be specified to limit over fitting. If max features are 0.5, only a random half of all the features are considered and evaluated for finding the best split.

V. Results and Conclusion

Efficient pricing of real estate customers for their priorities and budgets must therefore be predicted. This project efficiently analyzes past industry trends and price ranges to predict future prices.

Logistic Regression: Accuracy is constant between 0.45 to 0.55 threshold (Accuracy: 80.5-81.4)

Naïve Bayes: More towards predicting High Sensitivity

Decision Tree C5.0: Concentrated on Generating rules, Gives best results (Accuracy, TNR, TPR, above 92%)

Ada Boost: More towards predicting High Sensitivity

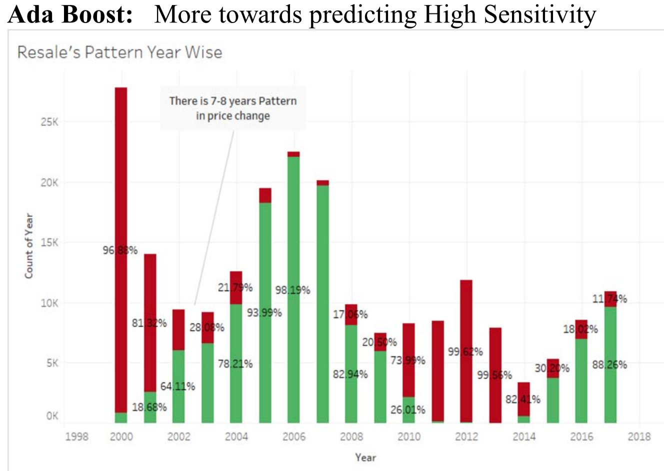


Fig. 2. Resale pattern analysis

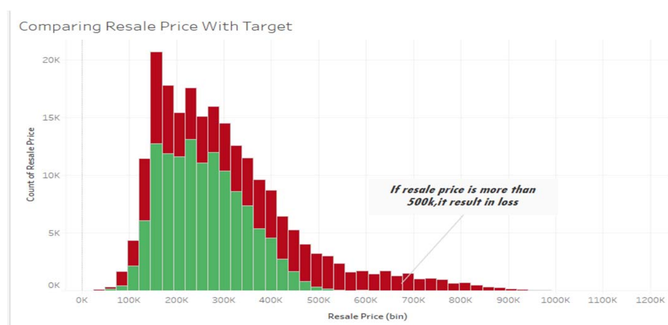


Fig. 3. Comparing the resale price with a target

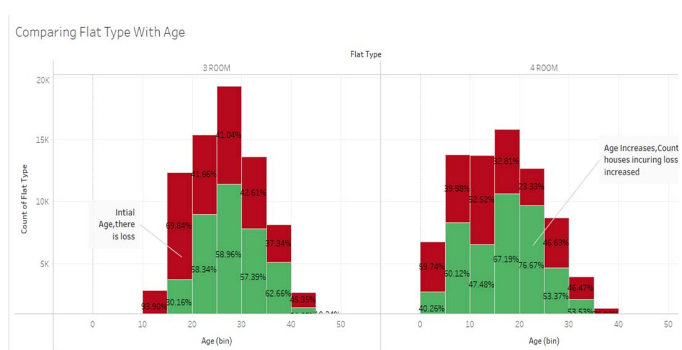
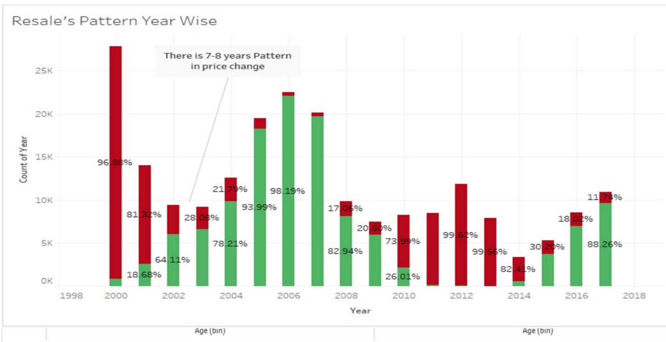


Fig. 4. Comparing flat type with age

978-1-7281-0027-2/19/$31.00 ©2019 IEEE

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on October 04,2024 at 04:11:55 UTC from IEEE Xplore. Restrictions apply.



Algorithm	Accuracy
Logistic regression	81.5%
Decision tree	92%
Random forest	86.5%
Naive bayes	88%
Ada boost	96%

Fig. 5. performance analysis of models

Finally, Adaboost and Decision Tree using C 5.0 is more concentrated on generating rules, it works well for this dataset. Using rules, we can predict the value of the house will result in profit or loss.

References

[1] Lu, Sifei, et al. "A hybrid regression technique for house prices prediction." Industrial Engineering and Engineering Management (IEEM), 2017 IEEE International Conference on. IEEE, 2017

[2] Mukhlishin, Muhammad Fahmi, Ragil Saputra, and Adi Wibowo. "Predicting house sale price using fuzzy logic and K-Nearest Neighbor." Informatics and Computational Sciences (ICICoS), 2017 1st International Conference on. IEEE, 2017

[3] Atharva chogle1, Priyanka khaire2, Akshata gaud3, Jinal Jain "House Price Forecasting using Data Mining Techniques", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 6, December 2017

[4] Febrita, Ruth Ema, et al. "Data-driven fuzzy rule extraction for housing price prediction in Malang, East Java." Advanced Computer Science and Information Systems (ICACSIS), 2017 International Conference on. IEEE, 2017

[5] Lim WT, Wang L, Wang Y, Chang Q. Housing price prediction using neural networks. In Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2016 12th International Conference on 2016 Aug 13 (pp. 518-522). IEEE.

[6] Tan F, Cheng C, Wei Z. Time-Aware Latent Hierarchical Model for Predicting House Prices. In2017 IEEE International Conference on Data Mining (ICDM) 2017 Nov 1 (pp. 1111-1116). IEEE.

[7] Banerjee D, Dutta S. Predicting the housing price direction using machine learning techniques. In 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI) 2017 Sep 21 (pp. 2998-3000). IEEE.

[8] Wang JJ, Hu SG, Zhan XT, Luo Q, Yu Q, Liu Z, Chen TP, Yin Y, Hosaka S, Liu Y. Predicting house price with a memristor-based artificial neural network. IEEE Access. 2018;June(6):(pp.16523-16528).IEEE.

[9] Febrita RE, Alfiyatin AN, Taufiq H, Mahmud WF. Data-driven fuzzy rule extraction for housing price prediction in Malang, East Java. In Advanced Computer Science and Information Systems (ICACSIS), 2017 International Conference on 2017 Oct 28 (pp. 351-358). IEEE.

[10] Yoon JH, Baldick R, Novoselac A. Dynamic demand response controller based on real-time retail price for residential buildings. IEEE Transactions on Smart Grid. 2014 Jan;5(1):(pp.121-9).IEEE.

[11] Abbasov C. The prediction of the chance of selling of houses as the factor of financial stability. InApplication of Information and Communication Technologies (AICT), 2016 IEEE 10th International Conference on 2016 Oct 12 (pp. 1-4). IEEE.

978-1-7281-0027-2/19/$31.00 ©2019 IEEE

Authorized licensed use limited to: Nanyang Technological University Library. Downloaded on October 04,2024 at 04:11:55 UTC from IEEE Xplore. Restrictions apply.