A Novel Approach for Predicting House Price Using K-Nearest Neighbors Algorithm Comparing Accuracy Prediction with Linear Regression



A NOVEL APPROACH FOR PREDICTING HOUSE PRICE USING K-NEAREST NEIGHBORS ALGORITHM COMPARING ACCURACY PREDICTION WITH LINEAR REGRESSION

Z.Sri Sai Swetha¹, R. Dhanalakshmi^{2*}

Article History: Received: 12.12.2022	Revised: 29.01.2023	Accepted: 15.03.2023

Abstract

Aim: This work is a general research of artificial intelligence K-Nearest Neighbors algorithm (KNN) and Novel linear regression (LR) for the house price prediction systems to chip away the exactness and the accuracy of house price prediction.

Materials and Methods: From Machine Learning, Novel K-Nearest Neighbors algorithm (N=10) and Novel linear regression(N=10) techniques are mimicked by varying the KNN parameter and Novel linear regression to propel the ph. The sample size was calculated using the G power of 80% for two groups and there are 20 samples used in this work.

Results: Based on the obtained results KNN has significantly better classification accuracy (67.92%) compared to the Novel Linear Regression (53.46).Statistical significance difference between long short term memory and Novel Artificial neural network was found to be 0.278 (p>0.05) which infers that both groups are insignificant. **Conclusion:** The K-Nearest Neighbors algorithm produces better results in prediction on house price monitoring to improve the price prediction accuracy than the Novel Linear Regression.

Keywords: Artificial Intelligence, Machine Learning, K-Nearest Neighbors, Novel Linear Regression, Prediction, Classification.

¹Research Scholar, Department of Computer Science and Engineering,Saveetha School of Engineering,Saveetha Institute of Medical and Technical Sciences,Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105. ^{2*}Project Guide, Department of Computer Science and Engineering,Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences,Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105.

1. Introduction

In this research work, artificial intelligence and machine learning is a good prediction method that had a great impact on house price prediction (Akyuz, Erdogan, and Yildiz, n.d.), which makes the price prediction based on many houses and their land zones in the city. The destruction of fake prices acts like perhaps the best danger to the idea of consistent truth using novel classification accuracy. Since the advancement of the fake prices in prediction of the house, there has been a developing joint exertion by the scholarly local area to explore and create the approaches fit for the investigation (Kirkeby and Larsen 2021a) and also used to recognize and intercede an exhibition of fraud in house price prediction. The importance of this research study is to identify the logical proof that has effectively uncovered the multiple fake prices of a house bogus actions in search of houses (Chai et al. 2018) the various artificial intelligence applications of this research study is real estate price prediction system and to apply limited features of house price predictions up to an arbitrary likelihood of roughly 62% correctness (Sarip and Hafez 2015). Moreover, the battle against the unregistered house dealers, the informal community and the plot utilization issues are indistinguishable.

The Literature survey in house price prediction systems and many comparisons have also been done for this prediction. Approximately 34 research articles published on house price prediction systems in IEEE xplore and 124 articles in google scholar and 25 articles were found in science direct. Although the accuracy of the price prediction software has remarkably developed and marked over the past two decades (Samukcham 2021), it is important to note that the automatic price prediction in house price prediction systems is not entirely automatic (Kirkeby and Larsen 2021b). Identification accuracy can be a bit poor in cases where the image captured conditions are not optimal (Kaushal and Shankar, n.d.). To manage this uncertainty, in many applications algorithms have human users with a 'house list' displaying the topmost matching images which were collected from the database ranked in the order of resemblance to the scrutiny image (Deghi et al. 2020).Our team has extensive knowledge and research experience that has translated into high quality publications(K. Mohan et al. 2022; Vivek et al. 2022; Sathish et al. 2022; Kotteeswaran et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Yaashikaa, Senthil Kumar, and Karishma 2022; Saravanan et al. 2022; Jayabal et al. 2022; Krishnan et al. 2022; Jayakodi et al. 2022; H. Mohan et al. 2022)

The research gap identified from the literature survey is that there are many methods proposed for house price prediction systems but some of the methods produce less accuracy rate. The study compares various machine learning algorithms accuracy in predicting house price prediction systems such as K-Nearest Neighbors algorithm, Novel Linear Regression, linear search vector machine algorithm. laboratory of neurogenetics and neuroscience algorithm. K-Nearest neighbors algorithm for house price prediction system. In this system, the algorithm tracks the false status of house's price in house price prediction (Yu and Gastwirth 2010) has implemented in accuracy for artificial intelligence in machine learning algorithm such as linear search vector machine algorithm, laboratory of neurogenetics and neuroscience algorithm, K-Nearest neighbor algorithm. The aim of the study is to show that the K-Nearest Neighbors algorithm appears to perform better among all the algorithms. In the previous study the efficiency of predicting the house price in a city by using classification of accuracy with the proposed K-Nearest Neighbors algorithm to improve the classification accuracy. Eventually, the proposed approach's validity is estimated utilizing data recovery measurements like affectability, exactness, and precision (Deghi et al. 2020).

2. Materials and Methods

The research work was carried out in the Programming Machine Learning, Lab of Department of Computer Science and Engineering at Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai. The sample size has been calculated using G power software by comparing the process and their results. In each group, Two numbers of groups are selected for comparing the process and their result. In each group, 20 sets of samples and 40 samples in total are selected for this work. The pre-test power values calculated using G power 3.1 software (g power setting the parameters: Statistical difference between two independent means, α=0.05, power=0.80, two algorithms (K-Nearest Neighbors and LR algorithm) are implemented using technical analysis software.In this work, no human and animal samples were used so no ethics are required (G and Sundar 2020).

K-Nearest Neighbors

The K-Nearest Neighbors is one in all the only machine learning algorithms which support supervised learning techniques. The K-NN algorithmic program assumes the similarity between the new data and market cases and places the new case into the class, that's the most almost like market classes. KNN algorithmic program stores all the on the market knowledge and classifies a replacement information supporting the similarity.

The main principle behind nearest methods is

- To find a predefined number of training samples close in distance to the new data point
- To predict the new labeo from these number of training samples

K-NN could be the non-parametric algorithmic program, which implies it doesn't build any assumption on underlying knowledge. It's conjointly referred to as a lazy learner algorithmic program as a result of it doesn't learn from the coaching set directly instead it stores the dataset and at the time of classification, it performs associate action on the dataset

This algorithm is used to take care of grouping model issues. The K-Nearest Neighbors algorithm or KNN algorithm essentially makes an imaginary boundary to classify the data. When new data points come in, the algorithm will try to anticipate that to the nearest to the boundary line.

Accordingly larger k value means smoother covers of separation resulting in less complicated models, though more modest k value tends to over fit the data and resulting in complicated models. Using the K nearest neighbor algorithm that fits the historical data and predicts the future.

Input: Training and Testing data

Output: Accuracy

- Step 1. Import dataset
- Step 2. Preprocess the data

Step 3. x,y (define the features needed)

Step 4. Split the dataset into train and test data

Step 5. Define the model

Step 6. Define the evaluation method

Step 7. Fit the model on the training and test data

Step 8. evaluate the model on the dataset

Step 9. Report performance

Novel Linear Regression Algorithm

The Novel Linear Regression algorithm is a machine learning algorithm from artificial intelligence which is based on supervised learning. This Novel linear regression performs regression tasks. Based on the independent variables regression models a goal prediction value. It is mostly utilized in predicting and determining the link between the variables. Various regression models vary in terms of the type of relationship they evaluate between dependent and independent variables are employed.

Novel Linear regression is used to predict the dependent variable (y) based on the independent variable (x). As a result of this regression approach

a linear connection between x (input) and y (output) is discovered (output).

The mathematical representation of Novel Linear Regression is

Y=a0+a1*X+ε

The hardware configuration was an Intel Core i5 processor with a RAM size of 8GB. The system type used was a 64bit OS, X64 based processor with an HDD of 1TB. The operating system used was Apple and the tool used for implementation was Jupyter Notebook with python programming language.

Input: Training and Testing data

Output: Accuracy

Step 1. Import dataset

Step 2. Preprocess the data

Step 3. x,y (define the features needed)

Step 4. Split the dataset into train and test data

Step 5. Define the model

Step 6. Define the evaluation method

Step 7. Fit the model on the training and test data

Step 8. evaluate the model on the dataset

Step 9. Report performance

Statistical Analysis

SPSS software is used for statistical analysis of KNN and the LR algorithm between house price predicting groups was done through SPSS Version 2. The independent variable is KNN accuracy and the dependent variable is efficiency. The independent T-test analyses are carried out to calculate the accuracy of the KNN for both methods (Kaushal and Shankar, n.d.). The proposed K- Nearest Neighbors algorithm approach gives accuracy 67.92% which is significantly better in classification accuracy when compared to the Novel Linear Regression which has less accuracy 53.46 %.

3. Results

Table 1 shows the simulation results of the proposed algorithm K-Nearest Neighbors and the existing system Novel linear regression were run at different intervals in the apple os and terminal with a sample size of 20. From Table 1, it was observed that the mean accuracy of the KNN algorithm was 67.92% and the Novel Linear Regression was 53.46%.

Table 2 represents the T-test comparison of both the KNN algorithm and LR algorithm. The mean, standard deviation and standard error mean were calculated by taking an independent variable T-test along with the Novel Linear Regression with a value of 0.156 and effect size =1.313.This also represents the Mean of the KNN algorithm which is better compared with the Novel linear regression algorithm with a standard deviation of 1.01192 and 945.43 respectively. From the results, the KNN algorithm (67.92%) gives better accuracy than the Novel Linear Regression (53.46%). The bar chart represented by Fig. 1 shows the comparison chart of KNN of Novel Linear Regressions in terms of mean and accuracy. The mean accuracy of the KNN algorithm is better than the Novel linear regression.

4. Discussion

The KNN and Novel Linear Regressions are implemented and compared the prices for prediction to improve the accuracy of house prices. From obtained results, it is concluded that the KNN algorithm provides better accuracy results compared to the Novel Linear Regression. The Data evaluation was performed using KNN SPSS tool (Kaushal and Shankar, n.d.). The experiments were conducted among the study groups K-Nearest Neighbors and linear regression by varying sample size.

In the recent surveys, the proposed (Kamencay et al. 2016; Costa and da Costa, n.d.) K-Nearest Neighbors has increased throughput value with 67.9200 and significance valueless with 0.039 (Kamencay et al. 2016). In this research work, Novel linear regressions and the K-Nearest Neighbors algorithms were analyzed for increasing the throughput percentage of student access privilege. It is observed that the Novel K-Nearest Neighbors appear to increase the throughput compared to the Novel Linear Regression (53.46). The main advantage of using KNN algorithms is fast processing, high quality of output. The disadvantage of using LR algorithms is less accuracy, more time complexity. The sigmoid function maps the predictions to the probabilities of the K-Nearest Neighbors based on the access privilege The experimental results of machine learning and artificial intelligence show evidence there is a significant difference between the Novel linear regression and K-Nearest Neighbors algorithms (Kaushal and Shankar, n.d.). In this paper, the Novel Linear Regression was implemented with an accuracy percentage of 53.46. From the above discussion, it was observed that the factors which affect the current research study were due to attributes that affect the throughput percentage of house price databases in the research work are title, text, subject, data (G and Sundar 2020). The attributes that mainly concentrated to increase the throughput were to predict the correct price of a house.

The limitations of the proposed work are one of the attributes in the price prediction dataset used for predicting the house price in the residential area. The easy access and exponential growth of the

information available on the residential database have made it intricate to distinguish between original and fake price. House price prediction has limited paper prediction ability based on future prices of the house, which makes better house price prediction in the future. The Novel Linear Regression can address future predictions of house prices.

5. Conclusion

The main aim of the study is to implement an efficient method to find the house price prediction. This research study applied the K-Nearest Neighbors algorithm using linear regression algorithm to find the house price prediction. The proposed KNN proved with better accuracy of 67.92% when compared to novel linear regression accuracy is 53.46 % for predicting the price of a house in a locality.

Declaration

Conflict of Interests

No conflicts of interest in these manuscripts

Authors Contribution

Author ZSSS was involved in data collection, data analysis, manuscript writing. Author RD was involved in conceptualization, data validation, and critical review of the manuscript.

Acknowledgments

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

- 1. Railsdata Software Pvt Ltd
- 2. Saveetha University
- 3. Saveetha Institute of Medical and Technical Sciences
- 4. Saveetha School of Engineering

6. References

- Akyuz, Sureyya, Birsen Eygi Erdogan, and Ozlem Yildiz. n.d. "A Novel Hybrid House Price Prediction Model." https://doi.org/10.22541/au.160284063.3790 9495/v1.
- Chai, Zhilei, Wei Song, Qinxin Bao, Feng Ding, and Fei Liu. 2018. "Taking Advantage of Hybrid Bioinspired Intelligent Algorithm with Decoupled Extended Kalman Filter for Optimizing Growing and Pruning Radial Basis Function Network." *Royal Society Open Science* 5 (9): 180529.

A Novel Approach for Predicting House Price Using K-Nearest Neighbors Algorithm Section A-Research paper Comparing Accuracy Prediction with Linear Regression

- Costa, Leandro Moreira da, and Leandro Moreira da Costa. n.d. "VOTING NEAREST NEIGHBORS: SVM CONSTRAINTS SELECTION ALGORITHM BASED ON K-NEAREST NEIGHBORS." https://doi.org/10.23860/diss-moreira-dacosta-leandro-2019.
- Deghi, Andrea, Mitsuru Katagiri, Mr Sohaib Shahid, and Nico Valckx. 2020. Predicting Downside Risks to House Prices and Macro-Financial Stability. International Monetary Fund.
- G, Sundar, and G. Sundar. 2020. "Improvement of Stock Price Prediction by Synthesizing Recurrent ANN Back Track Solver." *International Journal of Psychosocial Rehabilitation*. https://doi.org/10.37200/ijpr/v24i1/pr200297
- Jayabal, Ravikumar, Sekar Subramani, Damodharan Dillikannan. Yuvarajan Devarajan, Lakshmanan Thangavelu, Nedunchezhiyan, Mukilarasan Gopal Kaliyaperumal, and Melvin Victor De Poures. 2022. "Multi-Objective Optimization of Performance and Emission Characteristics of a CRDI Diesel Engine Fueled with Sapota Methyl Ester/diesel Blends." Energy. https://doi.org/10.1016/j.energy.2022.12370 9.
- Jayakodi, Santhoshkumar, Rajeshkumar Shanmugam, Bader O. Almutairi, Mikhlid H. Almutairi, Shahid Mahboob, M. R. Kavipriya, Ramesh Gandusekar, Marcello Nicoletti, and Marimuthu Govindarajan. 2022. "Azadirachta Indica-Wrapped Copper Oxide Nanoparticles as a Novel Functional Material in Cardiomyocyte Cells: An Ecotoxicity Assessment on the Embryonic Development of Danio Rerio." Environmental Research 212 (Pt A): 113153.
- Kamencay, Patrik, Tibor Trnovszky, Miroslav Benco, Robert Hudec, Peter Sykora, and Andrej Satnik. 2016. "Accurate Wild Animal Recognition Using PCA, LDA and LBPH." 2016 ELEKTRO. https://doi.org/10.1109/elektro.2016.751203 6.
- Kaushal, Anirudh, and Achyut Shankar. n.d. "House Price Prediction Using Multiple Linear Regression." SSRN Electronic Journal.

https://doi.org/10.2139/ssrn.3833734.

- Kirkeby, Sara J., and Vegard Høghaug Larsen. 2021a. House Price Prediction Using Daily News Data.
 - —. 2021b. House Price Prediction Using Daily News Data.

- Kotteeswaran, C., Indrajit Patra, Regonda Nagaraju, D. Sungeetha, Bapayya Naidu Kommula, Yousef Methkal Abd Algani, S. Murugavalli, and B. Kiran Bala. 2022. "Autonomous Detection of Malevolent Nodes Using Secure Heterogeneous Cluster Protocol." *Computers and Electrical Engineering*. https://doi.org/10.1016/j.compeleceng.2022. 107902.
- Krishnan, Anbarasu, Duraisami Dhamodharan, Thanigaivel Sundaram, Vickram Sundaram, and Hun-Soo Byun. 2022. "Computational Discovery of Novel Human LMTK3 Inhibitors by High Throughput Virtual Screening Using NCI Database." Korean Journal of Chemical Engineering. https://doi.org/10.1007/s11814-022-1120-5.
- Mohan, Harshavardhan, Sethumathavan Vadivel, Se-Won Lee, Jeong-Muk Lim, Nanh Lovanh, Yool-Jin Park, Taeho Shin, Kamala-Kannan Seralathan, and Byung-Taek Oh. 2022. "Improved Visible-Light-Driven Photocatalytic Removal of Bisphenol A Using V2O5/WO3 Decorated over Zeolite: Degradation Mechanism and Toxicity." *Environmental Research*. https://doi.org/10.1016/j.envres.2022.11313 6.
- Mohan, Kannan, Abirami Ramu Ganesan, P. N. Ezhilarasi, Kiran Kumar Kondamareddy, Durairaj Karthick Rajan, Palanivel Sathishkumar, Jayakumar Rajarajeswaran, and Lorenza Conterno. 2022. "Green and Eco-Friendly Approaches for the Extraction of Chitin and Chitosan: A Review." *Carbohydrate Polymers* 287 (July): 119349.
- Samukcham, Millee. 2021. "House Price Prediction Using Machine Learning." International Journal for Research in Applied Science and Engineering Technology. https://doi.org/10.22214/ijraset.2021.37451.
- Saravanan, A., P. Senthil Kumar, B. Ramesh, and S. Srinivasan. 2022. "Removal of Toxic Heavy Metals Using Genetically Engineered Microbes: Molecular Tools, Risk Assessment and Management Strategies." *Chemosphere* 298 (July): 134341.
- Sarip, Abdul G., and Muhammad Burhan Hafez. 2015. "Fuzzy Logic Application for House Price Prediction." *International Journal of Property Sciences.* https://doi.org/10.22452/ijps.vol5no1.3.
- Sathish, T., R. Saravanan, V. Vijayan, and S. Dinesh Kumar. 2022. "Investigations on Influences of MWCNT Composite Membranes in Oil Refineries Waste Water Treatment with Taguchi Route." *Chemosphere* 298 (July): 134265.

Vivek, J., T. Maridurai, K. Anton Savio Lewise, R.

A Novel Approach for Predicting House Price Using K-Nearest Neighbors Algorithm Section A-Research paper Comparing Accuracy Prediction with Linear Regression

Pandiyarajan, and K. Chandrasekaran. 2022. "Recast Layer Thickness and Residual Stress Analysis for EDD AA8011/h-BN/B4C Composites Using Cryogenically Treated SiC and CFRP Powder-Added Kerosene." *Arabian Journal for Science and Engineering*. https://doi.org/10.1007/s13369-022-06636-

https://doi.org/10.100//s13369-022-0663/ 5.

- Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity." *Fuel*. https://doi.org/10.1016/j.fuel.2022.123814.
- Yaashikaa, P. R., P. Senthil Kumar, and S. Karishma. 2022. "Review on Biopolymers and Composites – Evolving Material as Adsorbents in Removal of Environmental Pollutants." *Environmental Research*. https://doi.org/10.1016/j.envres.2022.11311 4.
- Yu, Binbing, and Joseph L. Gastwirth. 2010. "HOW WELL DO SELECTION MODELS PERFORM? ASSESSING THE ACURACY

OF ART AUCTION PRE-SALE ESTIMATES." *Statistica Sinica* 20 (2): 837–52.

- Beltratti, A. and Morana, C., "International house prices and macroeconomic fluctuations," Journal of Banking & Finance, pp.533-545, March 2010.
- Adams, Z. and Füss, R., "Macroeconomic determinants of international housing markets," Journal of Housing Economics, pp.38-50, March 2010.
- Otrok, C. and Terrones, M.E., "House prices, interest rates and macroeconomic fluctuations: international evidence," Unpublished manuscript.
- Mason, C. and Quigley, J.M., "Non-parametric hedonic housing prices," Housing studies, pp.373-385, July 1996.
- Tita, G.E., Petras, T.L. and Greenbaum, R.T., "Crime and residential choice: a neighborhood level analysis of the impact of crime on housing prices," Journal of quantitative criminology, p.299, December 2006.

TABLES AND FIGURES

Table 1. Group Statistics results Mean of K-Nearest Neighbors Algorithm is 67.92% is more compared toNovel Linear Regression 53.46% and standard error mean for KNN is 2.798 and LR is 2.697.

	Algorithms	N	Mean	Standard Deviation	Standard Error Mean
Accuracy	KNN	26	67.92	14.729	2.798
	LR	26	53.46	13.958	2.697

Table 2. This Independent sample t-test obtained the significance as 0.278 (p>0.05), significance, mean difference, std. error difference, and lower and upper interval difference. Independent samples t-test is applied for comparison of KNN and LR.

r								
		Levene's test for equality of variance	T-Test for Equality of Means					

A Novel Approach for Predicting House Price Using K-Nearest Neighbors Algorithm Section A-Research paper Comparing Accuracy Prediction with Linear Regression

		f	sig	t	df	sig	Difference	Std. Error Difference	95% Confide Interval Differer	nce Of nce
									Lower	upper
	Equal variance assumed	1.253	0.278	32.807	18	0.003	14.013	0.427	13.115	14.910
Accuracy	Equal Variance not assumed			32.807	17.734	0.003	14.013	0.427	13.114	14.910



Fig. 1. Comparison of K-Nearest Neighbors algorithm and Novel Linear Regression in terms of mean and accuracy. The mean accuracy of the KNN algorithm is better than the LR algorithm. X-axis: KNN vs LR, Y-axis: Mean accuracy of detection± 1SD