International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

HOUSING PRICE PREDICTION THROUGH BIG DATA ANALYSIS IN CNN USING MACHINE LEARNING

N. Seethalakshmi Karthikeyan*

* Civil Engineering, Rohini College of Engineering, Nagercoil

Abstract- Machine learning is a data analysis technique that automates the development of analytical models. It is a subfield of artificial intelligence that is founded on the premise that systems can learn from data, spot patterns, and make judgements. Housing prices have been one of the most concerning issues for individuals all over the world. The major goal of this research project is to forecast house prices using real-world factors. This study describes how to forecast housing costs using the CNN regression technique. Because home costs are inextricably linked to several criteria such as location, area, and car parking facilities, it takes all of this knowledge to estimate specific house prices.The proposed method takes into account the sophisticated factors employed in house price estimation and delivers a more accurate forecast. This article will also present an optimistic result for housing price projection for important cities in India, such as Chennai and Delhi.

Index Terms- Big Data Analysis, CNN, Machine learning

I. INTRODUCTION

Recently, machine learning (ML) has become very widespread in research and has been incorporated in a variety of applications, including text mining, spam detection, video recommendation, image classification, and multimedia concept retrieval. The continuing appearance of novel studies in the fields of machine learning is due to both the unpredictable growth in the ability to obtain data and the amazing progress made in the hardware technologies.

Urban real estate has consistently been not only a key part of household issue, but also a great collision or on copious economy. As such, it is of best moment to study the action of housing price, point to provide discernment in the real estate industry and help prepare genuine policy advice for the collaborator.

Earlier, it's a very popular and common practice to price the property without the proper evaluation of the land, infrastructure etc. We need a proper prediction on the real estate and the houses in housing market. We can see a mechanism that runs throughout the properties. Buying and selling a house will be a life time goal for most of the individual but there are a lot of people making mistake in India as most of the people are buying properties from the people, they don't know by seeing the news all around them. In India, people buy properties which are too expensive but it's not worth it. In the housing market 2022 the house sold in India was about 80 lakhs but the real price according locality and size was about 60 lakhs.

In earlier year, there was an economic collapse that give the clue to the impending disaster, this situation is currently happening and the prices of houses are getting higher compared to current economic situation of our country, the Indian government fails to produce the data about the houses so it was very difficult for peoples to buy the properties. Therefore, the people searched on internet for the evidence for house price.

II. LITERATURE REVIEWS

1. Estimation of Housing Prices by Fuzzy Regression and Artificial Neural Network

Reza Ghodsi et al: Changes in housing prices concern both individuals and government since they have substantial influence on the socio-economic conditions. Valuations of housing are necessary in order to assess the benefit and liabilities in housing section. The housing price in Iran is based on eight economic indices. The study of trends in housing price has been made by considering the related seasonal data from 16 years ago and using the techniques of Artificial Neural Network Back propagation (ANN-Back propagation) and Fuzzy regression. The results of our experiments indicate that the estimation error (Mean Absolute Percentage Error, "MAPE") in the ANN-Back propagation technique is less than that in Fuzzy regression. It can be shown, by comparing the estimated housing prices by applying the ANN technique with the observed ones, that the ANN technique has favorably estimated the trends in the changes of housing prices.

2. Housing price prediction using neural networks

Wan Teng Lim et al: The forecast of Singapore condominium prices is important for potential buyers to make informed decisions. This paper applies two algorithms to predict Singapore housing market and to compares the predictive performance of

artificial neural network (ANN) model, i.e., the multilayer perceptron, with autoregressive integrated moving average (ARIMA) model. The more superior model is used to predict the future condominium price index (CPI). The lower mean square error (MSE) of the ANN models showed the superiority of ANN over other predictive tools.

3. A Linear Regression Approach to Prediction of Stock Market Trading Volume: A Case

Farhad Soleimanian Gharehchopogh Et Al: Predicting daily behavior of stock market is a serious challenge for investors and corporate stockholders and it can help them to invest with more confident by taking risks and fluctuations into consideration. In this paper, by applying linear regression for predicting behavior of S&P 500 index, we prove that our proposed method has a similar and good performance in comparison to real volumes and the stockholders can invest confidentially based on that.

Predicting the stock market due to its importance and popularity among the masses and also small and large companies due to financial benefits and its low risk is a growing topic in research.Despite the risk of falling too much value per share due to market fluctuations rarely happens, but again, the risk is there. These fluctuations which effect on stock price and trading volume have some difficulties in predicting. The fluctuations effect on the behavior of people in terms of capital savings or investment, the stock price and the increase or decrease of risk for investors. Therefore, in general, predicting the stock market behavior through techniques and various methods is a useful tool to assist investors to act with greater certainty and taking the risks and volatility of an investment into consideration and know when to buy the cheapest price and when to sell to highest price.

III. SYSTEM REQUIREMENTS

Hardware Requirements

- CPU type : Intel Pentium 4
- Clock speed : 3.0 GHz
- Ram size : 8 GB
- Hard disk capacity : 500 GB
- Monitor type : 15 Inch color monitor

Software Requirement

- Operating System : Windows 10, Linux
- Language : Python
- Tool : Anaconda

Software Description

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.Python is a must for students and working professionals to become a great Software Engineer especially when they are working in Web Development Domain.

IV. METHODOLOGY

4.1 Data Pre Processing:

Data pre-processing is the term for operations on Data at the lowest level of abstraction. These operations do not increase Data information content but they decrease it if entropy is an information measure. The aim of pre-processing is an improvement of the Data that suppresses undesired distortions or enhances some Data features relevant for further processing and analysis task.

4.2 Data Segmentation

Data segmentation is the task of clustering parts of a Data together that belong to the same object class. This process is also called pixel-level classification. In other words, it involves partitioning Data into multiple segments or objects. Data segmentation results in more granular information about the shape of a Data and thus an extension of the concept of Object Detection. All picture elements or size belonging to the same category have a common label assigned to them. For example: Let's take a problem where the picture has to be provided as input for object detection.

4.3 Data Feature Extraction

Feature extraction is a part of the dimensionality reduction process, in which, an initial set of the raw data is divided and reduced to more manageable groups. So when you want to process it will be easier. The most important characteristic of these large data sets is that they have a large number of variables. These variables require a lot of computing resources to process. So Feature extraction helps to get the best feature from those big data sets by selecting and combining variables into features, thus, effectively reducing the amount of data**.**

4.4 Data Classification using CNN

Data classification is the task of assigning a label or class to an entire Data. Data are expected to have only one class for each Data. Data classification is perhaps the most important part of digital Data analysis. To classify the Data by using CNN deep Convolutional Neural Network Classification between objects is a complex task and therefore Data classification has been an important task within the field of computer vision. Based CNN algorithm is used to find the Data and shows the textual data for that particular Data and display the action. Our proposed algorithm gives the optimized result with high accuracy Data classification refers to the labelling of Data into one of a number of predefined classes. There are potentially n number of classes in which a given Data can be classified.

4.5 Data Prediction Result

Data Captioning is the task of describing the content of a Data in words. This task lies at the intersection of computer vision and natural language processing. Most Data captioning systems use an encoder-decoder framework, where an input Data is encoded into an intermediate representation of the information in the Data, and then decoded into a descriptive text sequence. Finally the action of the Data can converted to the text based on the action.

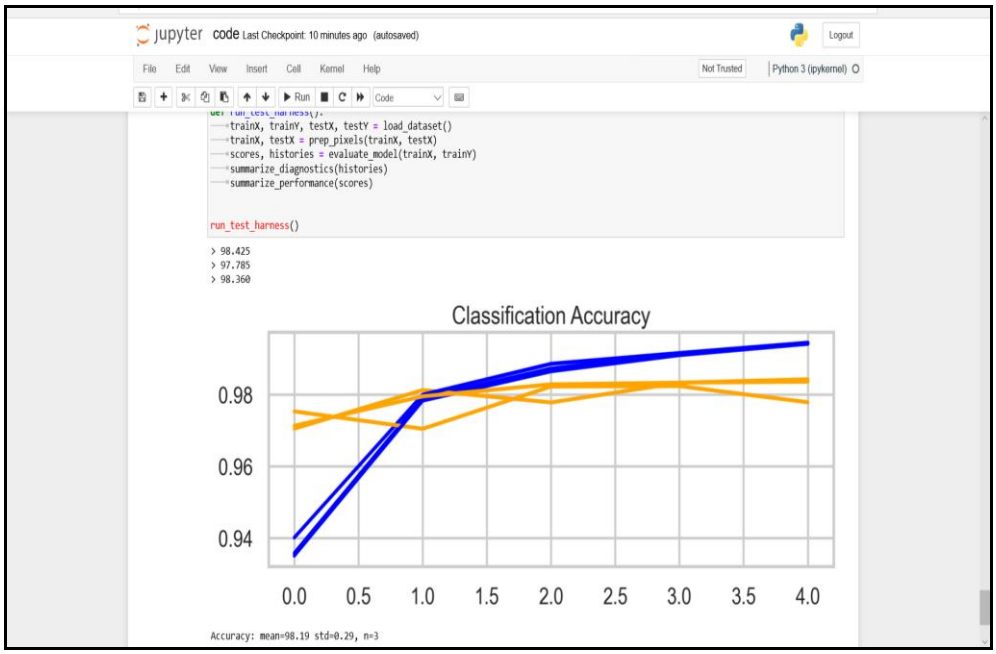
V. RESULTS

5.1 Sample Dataset

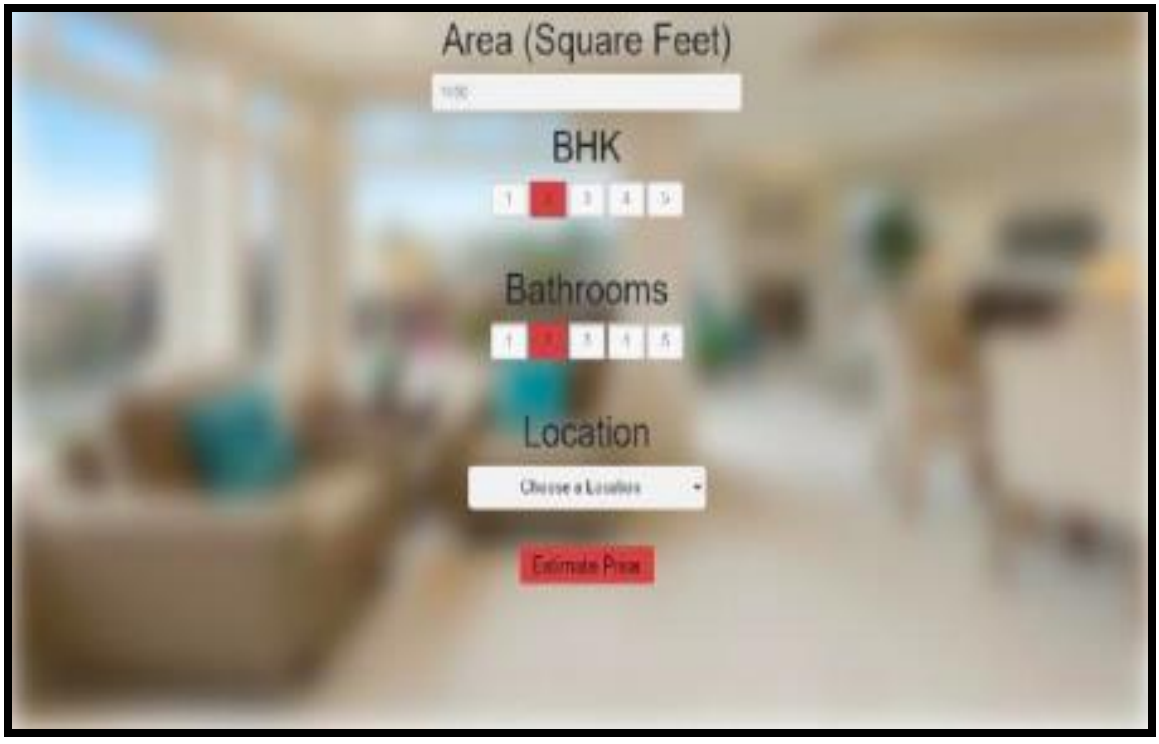
Posted On	BHK	Price	Size	House / Flat	Area Type	Area Locality	City	Furnishing Status	Car Parking	Bathroom	Point of Contact
24-04-2022	2	8000000	750	Individual House	Carpet Area	Rishi aurobindo sarani neat NIOH hospital	Kolkata	Semi-Furnished	No	2	Contact Owner
10-05-2022	2	7000000	685	Individual House	Super Area	Dum Dum Metro	Kolkata	Unfurnished	No	1	Contact Agent
24-05-2022	1	6500000	440	Individual House	Carpet Area	Baranagar	Kolkata	Furnished	No	1	Contact Owner
10-06-2022	4	13000000	1200	Individual House	Carpet Area	Tollygunge	Kolkata	Unfurnished	No	1	Contact Owner
23-06-2022	4	23040000	320	Flat	Carpet Area	Breach Candy	Mumbai	Furnished	Yes	4	Contact Agent
27-05-2022	5	26145000	450	Flat	Carpet Area	Bandra West	Mumbai	Furnished	Yes	5	Contact Agent
02-06-2022	1	12000000	225	Flat	Carpet Area	Ashok Van	Mumbai	Unfurnished	No	1	Contact Owner
02-06-2022	1	6500000	422	Flat	Carpet Area	Nalasopara West	Mumbai	Semi-Furnished	Yes	2	Contact Agent
10-05-2022	2	10500000	800	Individual House	Carpet Area	Ramamurthy Nagar	Bangalore	Semi-Furnished	No	2	Contact Owner
04-05-2022	1	8000000	350	Flat	Super Area	Thanisandra	Bangalore	Semi-Furnishe	Yes	1	Contact Owner

								d			
10-05-2022	1	9500000	900	Individual House	Super Area	J P Nagar	Bangalore	Semi-Furnished	No	1	Contact Owner
09-05-2022	2	13000000	700	Flat	Super Area	Vinayakanagar, Hebbal	Bangalore	Semi-Furnished	Yes	2	Contact Owner
04-06-2022	2	12000000	1000	Flat	Super Area	Laxmi Nagar	Delhi	Unfurnished	Yes	2	Contact Owner
14-05-2022	1	9000000	650	Flat	Super Area	Lado Sarai	Delhi	Unfurnished	Yes	1	Contact Owner
05-05-2022	2	14000000	750	Individual House	Super Area	Pandav Nagar	Delhi	Unfurnished	No	1	Contact Owner
10-05-2022	2	12500000	900	Flat	Super Area	Surya Apartment pocket a11	Delhi	Unfurnished	Yes	2	Contact Owner
12-05-2022	2	8000000	800	Flat	Super Area	Kayarambedu	Chennai	Semi-Furnished	Yes	2	Contact Owner
22-06-2022	1	7500000	560	Flat	Super Area	Thambaram west	Chennai	Unfurnished	Yes	1	Contact Owner
Posted On	**BHK**	**Price**	**Size**	**House / Flat**	**Area Type**	**Area Locality**	**City**	**Furnishing Status**	**Car Parking**	**Bathroom**	**Point of Contact**
13-05-2022	1	14000000	420	Individual House	Carpet Area	Mambalam West	Chennai	Unfurnished	Yes	1	Contact Owner
22-06-2022	3	12000000	930	Individual House	Super Area	Kelambakkam	Chennai	Semi-Furnished	Yes	2	Contact Owner
12-06-2022	2	12000000	1000	Individual House	Super Area	Ram nagar	Hyderabad	Unfurnished	No	2	Contact Owner
26-05-2022	1	8000000	550	Flat	Super Area	Rasoolpura, Mahatma Gandhi Road	Hyderabad	Semi-Furnished	Yes	1	Contact Owner
30-06-2022	2	12000000	933	Flat	Super Area	Kompally	Hyderabad	Unfurnished	Yes	2	Contact Owner
18-06-2022	2	10000000	1030	Flat	Super Area	Dr A.S. Rao Nagar	Hyderabad	Semi-Furnished	Yes	2	Contact Owner

5.2 Accuracy



5.3 Output



VI. CONCLUSION

Most models exhibited similar performance except for Multilayer Perceptron (MLP). As this model has several parameters to set, a reasonable approach would be to do an extensive exploration of best values. As this can be a tedious and error-prone task, one can make use of some automated search or evaluation process like the Grid Search resource available in the sci-kit-learn package.

With the use of a variety of analytical and graphical tools, we were able to evaluate the predictive performance of various housing price models applied to real data on single-family homes in major cities such as Chennai, Delhi,etc., in India. In addition, our models also helped identifying which characteristics of housing were most strongly associated with price and could explain most of the price variation. Furthermore, we were able to improve our models' prediction accuracy by accounting for the impact of spatial location.

We were able to identify most of the residential areas. There may be some more places that have housing complexes or multi-story apartments that are located in commercial areas. Such apartments were not included in this paper and can be counted in the future to give a more accurate result. With more and more demand for housing in metropolitan cities, there is a definite increase in the number of private builders that provide real estate with additional amenities to attract more customers.

There are several other models available that can be implemented for prediction. Data given as input to such a model should be compatible with the tool used and the operators involved in the process. Also, increasing the number of data sets can be used to increase the accuracy of the model.

ACKNOWLEDGEMENT

I express my thanks to **Prof. Dr. SAHAYA RUBEN., ME., Ph.D.,** Head of Civil Engineering Department, Rohini College of Engineering and Technology for his valuable advice and encouragement in completing the project. I express my hearty thanks to **Mr. RAJIV GANDHI,** Assistant Professor in Civil Department for his valuable guidance and encouragement in successfully completing the project.

REFERENCES

[1] G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

[2] W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[3] H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.

[4] B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.

[5] E. H. Miller, "A note on reflector arrays (Periodical style—Accepted for publication)," *IEEE Trans. Antennas Propagat.*, to be published.

[6] J. Wang, "Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication)," *IEEE J. Quantum Electron.*, submitted for publication.

AUTHOR

N. Seethalakshmi Karthikeyan, PG Student, Rohini College of Engineering, Nagercoil and cepriya92@gmail.com