www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

HOUSE PRICE PREDICTION USING MIXED DATA NEURAL NETWORKS

B. Ramakrishna Reddy[1], Depatment Of Mathematics

Mallikarjuna Rao Gundavarapu, B.Charith Chandra Sai, T.Jeevan Reddy, P.Sai Sree Venkat, Y, Karthik Reddy[2-6] , Department Of CSE

[1-6]Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India.

ABSTRACT:

House costs increment consistently, so there is a requirement for a framework to anticipate house costs from here on out. House value expectation can assist the designer with deciding the selling cost of a house and can assist the client with orchestrating the perfect opportunity to buy a house. There are three factors that impact the cost of a house which incorporate states of being, idea and area. States of being are properties moved by a house that can be seen by human detects, including the size of the house, the quantity of rooms, the accessibility of kitchen and carport, the accessibility of the nursery, the area of land and structures, and the age of the house. The goal of the endeavor is to predict the capable house assessing for land clients concerning their monetary plans and needs. Land is the most un-straight industry in our current circumstance. Dwelling costs keep on changing all week long and every so often are publicized instead of being established on valuation. Expecting to stop costs with real factors is the principal center of our investigation paper. Here we mean to make our appraisals subject to every fundamental limit that is considered while choosing the expense. We use different backslide systems in this pathway, and our results are not sole affirmation of one strategy rather it is the weighted mean of various techniques to give most accurate results. The results exhibited that this technique yields least bungle and most noteworthy precision than solitary estimations applied.

Keywords: *House rent costs, Machine learning,backslide systems, property place*

1. INTRODUCTION

Individuals looking to purchase another home will, on average, stick to more standard financial plans and market techniques. The current plan aims to assess the value of residential real estate without the need for critical forecasts of future market trends and price increases. The mission's objective is to anticipate talented residence estimation for land customers.in terms of their financial objectives and requirements Future costs can be predicted by looking outside market styles and price ranges, as well as papered improvements. This challenge's operation comprises

of a domain that identifies the client's decisions and then consolidates them utilizing direct relapse computation of data mining. In our environment, land is the most untrustworthy enterprise. The cost of lodging continues to rise each day of the week, and it is occasionally promoted rather than valued. Our research paper's major focus is on predicting house costs using real-world factors. While determining the cost, we recommend basing our tests on each critical limit. On this route, we utilize unique relapse tactics, and our outcomes aren't a one-way guarantee; rather, they're a weighted mix of numerous ways that provide the highest genuine results

Rationale

The general presentation could be estimated after foreseeing home charges for the explanation that forecast in numerous relapse calculations is predicated now at this point not best on a specific trademark anyway on an obscure assortment of traits that achieve the cost to be anticipated. House expenses rely on a man home detail. Houses have an adaptation assortment of capacities that probably won't have the indistinguishable worth because of its area. For example, an enormous home could likewise moreover have a superior expense assuming that it's far set in relevant rich area than being situated in a horrible area. A few Machine Learning calculations are utilized to cure issues inside side the genuine worldwide today. However, a number of them provide higher overall performance in positive circumstances, as said within side the No Free Lunch Theorem. Hence, this proposition attempts to apply relapse calculations and convolutional brain organization (CNN) to assess their general presentation as far as anticipating upsides of a given dataset.

GOAL:

The objective of our work is:

- As a client, we have evaluated specific parts of the ongoing structure that roused us to advance this application.
- The house cost forecast model assists the purchaser with picking the house in light of their cash status and furthermore saves the hour of house examination. This assists the dealer with bettering comprehend the worth of their property and gets them far from specialists.

OBJECTIVE:

Typically, the home value forecast has been physically created, which is exceptionally tedious and regularly leads to human error, and there are numerous weaknesses related to the sources of value that actually do not cover a large part of the cost of searching. The old technologies that people bought and sold no longer exist today. In the light of exploration, there are numerous agreements to address this issue. As the research journal shows, the recent general currency emergency has brought back a strong buzz in academic and procedural circles for resource costs and housing costs that are explicitly insured

near a money-related development. Also with regard to the demarcation with the statement "Housing are the advantages of the economic cycle" the truth is serious.

2. METHODOLOGY:

MACHINE LEARNING

The term AI alludes to the automated location of huge models in the data. In the beyond couple of years, this is not really a typical device for any organization, since it necessities to remove information from a lot of data. We're encircled by an AI-controlled advancement: the quantity of web records that will introduce us the best outcomes, spam-unfriendly programming, how we channel our messages, and Visa trades accomplished through an item that finds how to recognize con artists. to recognize faces and to find astute applications for individual assistance on cutting edge cell phones, how to see voice orders. The vehicles are furnished with crash aversion approaches that are collected utilizing man-made reasoning estimations. Artificial intelligence is a utilization of man-made thinking (AI) that empowers casings to normally absorb and work on a reality without being obviously changed. Backslide is a man-made consciousness instrument that creates assumptions by utilizing current, quantifiable data to acquire the connections between the target and other limits.

CONVOLUTIONAL NEURAL NETWORKS:

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics

RECURRENT NEURAL NETWORKS:

A recurrent neural network (RNN) is a special type of an artificial neural network adapted to work for time series data or data that involves sequences. Ordinary feed forward neural networks are only meant for data points, which are independent of each other. However, if we have data in a sequence such that one data point depends upon the previous data point, we need to modify the neural network to incorporate the dependencies between these data points. RNNs have the concept of ‘memory’ that helps them store the states or information of previous inputs to generate the next output of the sequence.

LOGISTIC REGRESSION:

Logistic regression is a process of modeling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome; something that can take two values such as true/false, yes/no, and so on. Multinomial logistic regression can model scenarios where there are more than two possible discrete outcomes. Logistic regression is a useful analysis method for

classification problems, where you are trying to determine if a new sample fits best into a category. As aspects of cyber security are classification problems, such as attack detection, logistic regression is a useful analytic technique.

MIXED DATA NEURAL NETWORKS:

Mixed data approach includes working with various kinds of free information. Autonomous information comprises of house pictures. House pictures are taken to an interesting size with the goal that it tends to be pre-handled. Also, compared without any problem. Convolutional Neural organizations are utilized for the house pictures for the structure of the model. Numeric information of the houses is pre-handled and examined utilizing relapse procedures. Regression analysis of mathematical information includes numerical conditions where information ought to be standardized prior to breaking down it. The layers of the Convolutional Neural networks and Regression analysis are linked and connected to form a complete connected layer

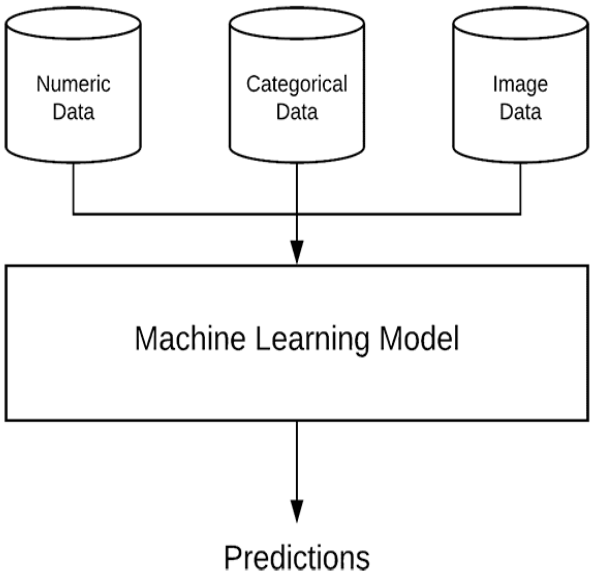

Fig 1. Workflow of mixed data neural networks

PYTHON:

Python is a significant level, interpreter-based programming language that might be utilized for an assortment of purposes. Guido van Rossum planned it and it was first conveyed in 1991. The Python plan's reasoning emphasis the code's rationality through its uncommon utilization of enormous blank area. Its language and item requested method were made to help programmers recorded as a hard copy steady code for little and enormous ventures.

Python assembles gradually, and trash is gathered. Procedural, object-requested, and utility writing computer programs are among the programming norms it supervises. In light of its immense standard library, Python is much of the time alluded to as the "batteries included" language.

ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

In the final part of the 1980s, Python was delivered as a trade for the ABC programming language. Python 2.0, which was sent off in 2000, included highlights like rundown understandings and a trash variety engineering for gathering reference cycles. Python 3.0, which was distributed in 2008, was a critical improvement to the language however, then again, doesn't work totally, and a large part of the Python 2 code doesn't run in Python

Python 2 (for instance, Python 2.7.x) will be censured on January 1, 2020 (introductory insurances taken in 2015), after which no security refreshes or different overhauls will be given. Python 2 will be expostulated, and just Python 3.5.x and higher will be upheld. A few systems have Python interpreters accessible. Python, an open source reference run, is made and kept up with by a worldwide local area of programmers. The Python Software Foundation is a non-benefit association that regulates and facilitates assets for Python and Python improvement.

SCIKIT-LEARN

In Python, Scikitlearn is perhaps the most helpful AI library. Arrangement, fallback, clustering, and dimensional loss are just a few of the AI and quantifiable visualization devices available in the sklearn library.

Scikitlearn provides a consistent Python interface for a variety of supervised and unsupervised learning techniques.

It is distributed on many Linux distributions and has a permissive simplified BSD licence, enabling academic and commercial use.

SciPy (Scientific Python) is the library's foundation, and it must be installed before scikitlearn can be used. This stack contains:

NumPy: Basic package for n-dimensional arrays SciPy: Fundamental Library for Scientific Computing Matplotlib: Comprehensive 2D / 3D Plot

IPython: Improved Interactive Console Sympy: Symbolic Math

Pandas: Data Structures and Analysis

Extensions or modules for SciPy maintenance, referred to as SciKits in the past. As a result, the module is known as scikitlearn and contains learning algorithms. The library's objective is to provide a level of reliability and support that is necessary for use in production systems. This entails a strong attention on usability, code quality, collaboration, documentation, and performance, among other things. Although the interface is Python, speed is improved by using additional libraries such as numpy for arrays and array operations, LAPACK, LibSVM, and Cpython.

NUMPY:

"Mathematical Python" is the name of a Python package called NumPy. It's the main library for the

logical register, which includes a fantastic n-dimensional display object; it allows devices to use C, C++, and other programming languages. It can also help with straight polynomial math, irregular number ability, and other subjects. NumPy is a Python math library that is available as open-source paper. A multidimensional display and frame information structures are included in NumPy. It's utilized to run a variety of numerical procedures, including geometric, factual, and arithmetic applications.

NumPy is a Numeric and Numarray extension.

NumPy is the finest Python library for logic processing. It includes, among other things, the following:

- An incredible N-dimensional cluster object
- Severe (transfer) capabilities

Tools for coordinating C / C ++ and Fortran code based on variables math, Fourier shift and irregular numerical capabilities NumPy can also be utilized as an effective multi-dimensional container of non-exclusive information for obvious logical applications. a wide range of data sources.

PANDAS:

Pandas is a collection of data manipulation and analysis software libraries built in Python. Its data structures and methods for manipulating numerical tables and time series are particularly useful. It's free software distributed under the BSD license, which has three clauses. The name comes from the word "panel data," which is an econometric term for data sets that contain observations for the same person over multiple time periods.

Library Features:

Data Frame object with inbuilt indexing for data manipulation

Data reading and writing tools for archaic data structures and various file formats.

Data synchronization and integrated missing data management.

Data set reorganization and pivoting

Segmentation based on tags, advanced indexing, and subsets of enormous data sets

Tag-based segmentation, complex indexing, and subsets of massive data sets are among the features available.

Split, apply, and combine operations on records using grouping by engine. Data records are merged and merged.

Hierarchical axis indexing for working with high-dimensional data in a low-dimensional data structure.

Time series functionality: date range generation and frequency conversion, rolling window statistics, rolling window linear regression, date shifting and delaying.

HTML:

HTML stands for Hypertext Markup Language, and it is a markup language that is used to display content in Web browsers. Cascading Style Sheets (CSS) and programming languages like JavaScript can help support it. HTML documents are retrieved from a Web server or local storage and converted into multimedia web pages by the Web browser. HTML is a semantic markup language that explains the structure of a website and provides the key to the document's appearance. HTML elements are the components that make up HTML pages. Images and other objects, such as interactive forms, can be embedded in displayed pages using HTML structure.HTML represents the structural semantics of text, such as headings, paragraphs, lists, links, quotations, and other elements, to construct structured documents. Tags enclosed in angle brackets are used to divide HTML elements.

HTML can be included with scripting languages (such as JavaScript) to influence the behavior and content of web pages.

INNOVATIVENESS:

Mixed data neural networks is the cutting edge approach and one of its sort. A few methodologies have been proposed at the house cost forecast utilizing machine learning, linear regression, data mining however they are restricted to predetermined number of boundaries. As this cutoff to less exactness and mistaken expectations. There ought to be a technique which can precisely anticipate the costs of the house by thinking about every one of the boundaries into thought. Mixed data neural networks approach is one such where various kinds of information can be combined, for example, images, numerical data and so forth to create precise outcomes.

USEFULNESS:

In our natural framework, land is the most unstable industry. Lodging costs vary consistently and are posted consistently as opposed to being founded on forecasts. Lodging costs estimating with genuine factors is key to our review paper. We intend to put together our assessments with respect to each material cutoff that is considered in settling on cost choices.

ACTIVITY DIAGRAM (User and Admin)

The activity diagram is a graphical depiction of the interaction flow inside a given situation. It's comparable to a flowchart in that it depicts the different actions that may be carried out in the system.

3. IMPLEMENTATION:

House price prediction using mixed data neural networks is carried out using various kinds of information. In Machine Learning, categorical data and image data. Individual information are fabricated utilizing different machine leaning algorithms. mixed data alludes to the idea of having numerous sorts of Independent data. We consider informational collection of various houses with

various boundaries. We would have numerous sorts of info information for a given House, including:

Numeric/continuous values, such as Area of house, Number of bedrooms, etc.

Categorical values, including Locality, City, etc.

Image data, such as any House images etc.

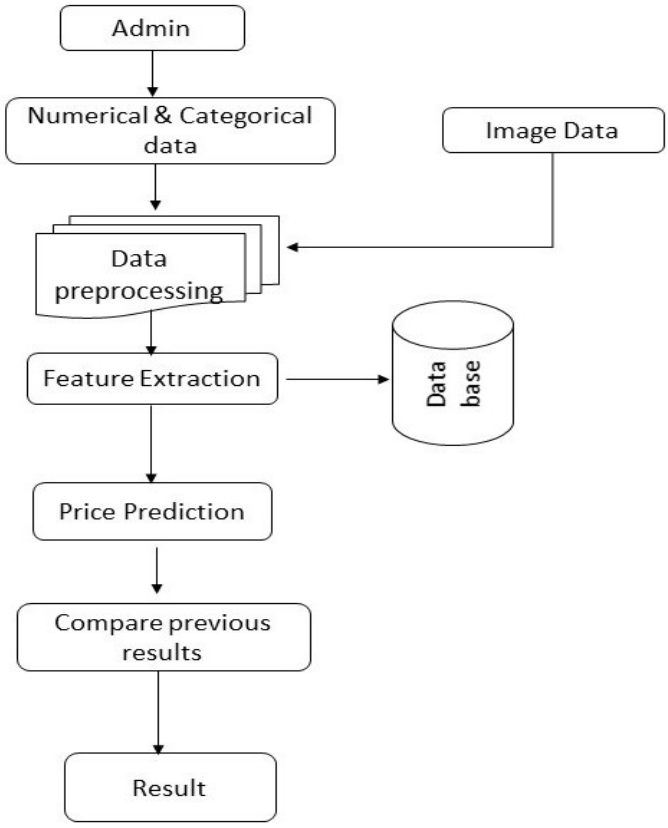

Fig-2 shows the Activity Diagram of our paper

OUR DATASET:

Our Approaches comprises of dataset which contains of in excess of 1300 mathematical information of different places where complete square feet of the house, number of overhangs, number of rooms and sort of society it is arranged in. The house information contained a few NULL qualities which can cause mistake in anticipating the result so we have standardized and adjusted the dataset in such a manner no NULL or rehashed values stays in the dataset.

Dataset is the most significant for any expectation model as a more noteworthy number of boundaries and values help in getting more precision and it will be reliable.

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view
count	21613	21613	21613	21613	21613	21613	21613	21613
mean	540088	3	2	2079	15106	1	0	0
std	367127	0	0	918	41420	0	0	0
min	75000	0	0	290	520	1	0	0
25%	321950	3	1	1427	5040	1	0	0
50%	450000	3	2	1910	7618	1	0	0
75%	645000	4	2	2550	10688	2	0	0
max	7700000	33	8	13540	1651359	3	1	4

condition	grade	sqft_above	sqft_basement	yr_built	yr_renovated	zipcode	lat	long
21613	21613	21613	21613	21613	21613	21613	21613	21613
3	7	1788	291	1971	84	98077	47	-122
0	1	828	442	29	401	53	0	0
1	1	290	0	1900	0	98001	47	-122
3	7	1190	0	1951	0	98033	47	-122
3	7	1560	0	1975	0	98065	47	-122
4	8	2210	560	1997	0	98118	47	-122
5	13	9410	4820	2015	2015	98199	47	-121

Fig.3 Different attributes of the dataset.

4. CONCLUSION

The picture information comprises of 4 kinds of pictures for every house which are washroom, room, kitchen and front facing which are standardized with the end goal that convolutional brain networks are applied on these pictures to create a model. Resize of pictures is finished by resize work, subsequent to resizing the pictures ought to be changed over into an exhibit by np. asarray() and put away into the meta dataset . Dataset ought to be Balanced with equivalent number of s skin cancer images.

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

Fig.4 Images of House dataset

The images are trained using Convolutional neural networks with one input layer ,2 hidden layers and one output layer. The model undergoes through 2 hidden layers in which max pooling and padding are applied to select and maximize the pixels, Activation function and is put at the end of neural networks that helps us to decide whether the neuron would fire or not. The model should be compiled and fitted so that it can be stored to compare input image.Fig.5 shows the implementation of CNN algorithm.

After creation of MLP and CNN models, the outputs of both the models are concatenated and connected to form a fully connected layer with four neurons to the combined input. The combined input to the final layers in the network is based on output of both MLP and CNN branches.

USER INTERFACE:

Initially the model is converted to a pickle file. A user interface is build using HTML, CSS and JAVA SCRIPT, where a back end server consists of all the information regarding the web page. we can give the input to the front end and generate the output using the pickle file from the back end. Estimate price button of the user interface gives the price of house based on the provided values. Prediction of the house price requires all the parameters as inputs for exact and perfect estimation.The user interface is simple and easy to use.

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

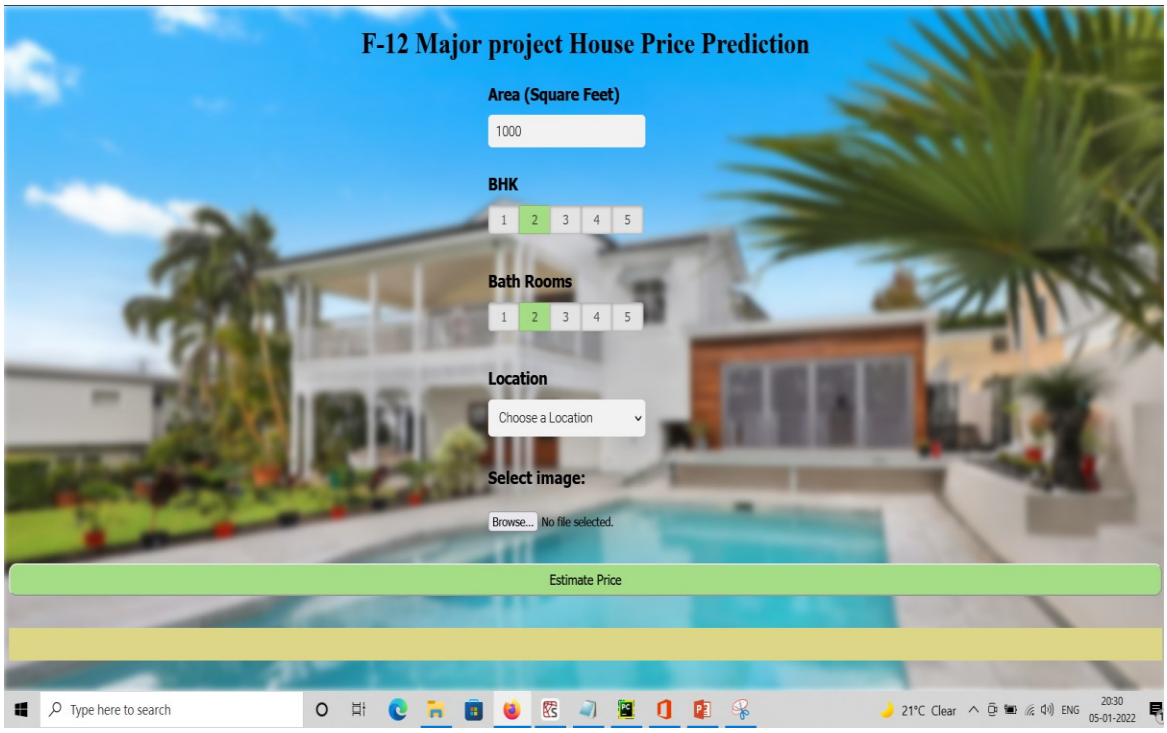

Fig. 5 User Interface

REFERENCES:

1. The DanhPhan, "Housing Price Prediction using Machine Learning Algorithms: The Case of Melbourne City, Australia (IEEE Paper)".
2. RushabSawant, YashwantJangidTushar Tiwari, Saurabh Jain and Ankita Gupta, "Comprehensive Analysis of Housing Price Prediction in Pune using Multi-Featured Random Forest Approach (IEEE Paper)".
3. http://www.99acres.com/property-rates-and-pricetrendsin-mumbai
4. The DanhPhan, "Housing Price Prediction using Machine Learning Algorithms: The Case of Melbourne City, Australia (IEEE Paper)"
5. Iain Pardoe, 2008, Modelling Home Prices Using Realtor Data
6. https://machinelearningmastery.com/machine-learning-with-python/
7. https://www.python.org/
8. https://medium.com/analytics-vidhya/deploying-a-machine-learning-model-on-web- using-flask-and-python-54b86c44e14a