



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Special Issue 2, April 2023

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Prediction of House Price Using SMLT

Sruthi nath <sup>1</sup>, Nithi Sneha <sup>2</sup>, Vandana S<sup>3</sup>

Associate Professor, Department of Computer Engineering, Velammal Engineering College, Chennai, Tamil Nadu, India <sup>1</sup>

U.G. Student, Department of Computer Engineering, Velammal Engineering College, Chennai, Tamil Nadu, India <sup>2-3</sup>

**ABSTRACT:** Generally, predicting how the house price will perform is one of the most difficult things to do. It can be described as one of the most critical process to predict that. This is a very complex task and has uncertainties. To prevent this problem in One of the most interesting (or perhaps most profitable) time series data using machine learning techniques. Hence, house price prediction has become an important research area. The aim is to predict machine learning based techniques for house price prediction results in error based calculation. The analysis of dataset by supervised machine learning technique(SMLT) to capture several information's like, variable identification, uni-variate analysis, bi-variant and multi-variant analysis, missing value treatments and analyze the data validation, data cleaning/preparing and data visualization will be done on the entire given dataset. To propose a machine learning- based method to predict the house price Index value by prediction results in the form of house price increase or stable state best regression from comparing supervise ML algorithms. Additionally, to compare and discuss the performance of various ML algorithms. dataset with evaluation classification report, identify the confusion matrix and the result shows that the effectiveness of the proposed ML algorithm technique can be compared with best accuracy MAE,MSE,R2

**KEYWORDS:** DenseNet, Mean Absolute Error, Regression.

## I. INTRODUCTION

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains. Data science can be defined as a blend of mathematics, business acumen, tools, algorithms and machine learning techniques, all of which help us in finding out the hidden insights or patterns from raw data which can be of major use in the formation of big business decisions.

## II. RELATED WORK

Urban housing price is widely accepted as an economic indicator which is of both business and research interest in urban computing. However, due to the complex nature of influencing factors and the sparse property of transaction records, to implement such a model is still challenging. To address these challenges, in this work, we study an effective and fine-grained model for urban subregion housing price predictions. Compared to existing works, our proposal improves the forecasting granularity from city-level to mile-level, with only publicly released transaction data[1]. We employ a feature selection mechanism to select more relevant features. a fine-grained forecasting model, JGC MMN, for subregion spatiotemporal housing price prediction. In particular, we modify the structure of DenseNet and adopt the method of bagging by fusing the KF-based method to improve the accuracy.

## III. METHODOLOGY

**Data Pre-processing :** Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered as close to the true error rate of the dataset. If the data volume is large enough to be representative of the population, you may not need the validation techniques. However, in real-



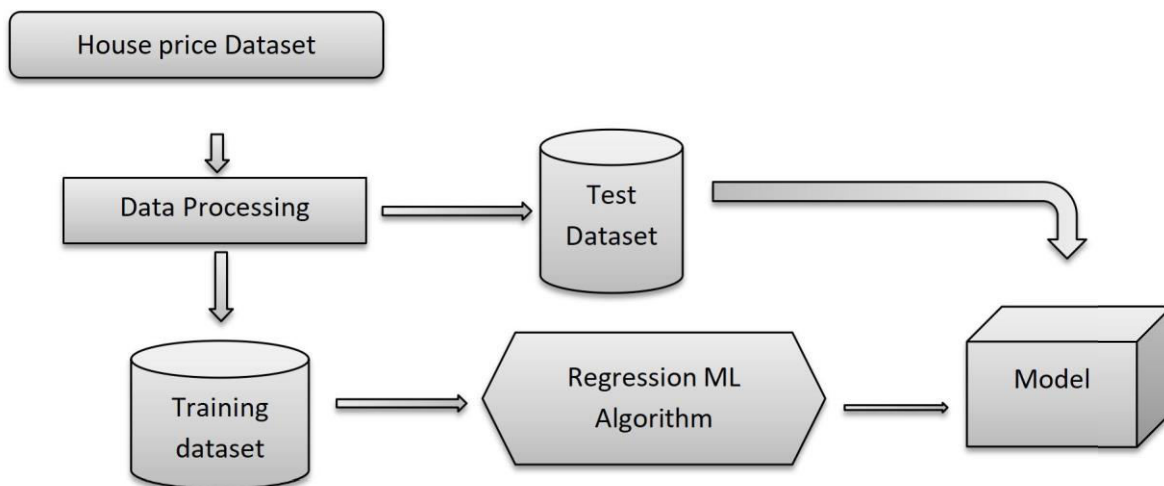
world scenarios, to work with samples of data that may not be a true representative of the population of given dataset. To finding the missing value, duplicate value and description of data type whether it is float variable or integer. The sample of data used to provide an unbiased evaluation of a model fit Data Processing Test dataset Training dataset Regression ML Algorithm Model on the training dataset while tuning model hyper parameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration. The validation set is used to evaluate a given model, but this is for frequent evaluation. It as machine learning engineers use this data to fine-tune the model hyper parameters[2]. Data collection, data analysis, and the process of addressing data content, quality, and structure can add up to a time- consuming to-do list. During the process of data identification, it helps to understand your data and its properties; this knowledge will help you choose which algorithm to use to build your model.

**Data Validation/ Cleaning/Preparing Process :** Importing the library packages with loading given dataset. To analyzing the variable identification by data shape, data type and evaluating the missing values, duplicate values. A validation dataset is a sample of data held back from training your model that is used to give an estimate of model skill while tuning models and procedures that you can use to make the best use of validation and test datasets when evaluating your models[3].

**Exploratory Data Analysis of House Prediction :** Multiple datasets from different sources would be combined to form a generalized dataset, and then different machine learning algorithms would be applied to extract patterns and to obtain results with MAE[4].

**Building the Regression model :** The predicting the house price problem, decision tree regression model is effective because of the following reasons: It provides better results in regression problem.

- It is strong in preprocessing outliers, irrelevant variables, and a mix of continuous, regression and discrete variables[4].
- It produces out of bag estimate error which has proven to be unbiased in many tests and it is relatively easy to tune with.



1(a) Architecture Of Proposed Model

Fig 1(a), The system architecture shown above was created to show the architecture of the proposed model.

**ALGORITHM AND TECHNIQUES**

**Algorithm Explanation :** In machine learning and statistics, classification is a supervised learning approach in which the computer program learns from the data input given to it and then uses this learning to classify new observation. This data set may simply be bi-class (like identifying whether the person is male or female or that the mail is spam or non-spam) or it may be multi-class too[4].



**Random Forest Regression :** Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model. After importing the libraries, importing the dataset, addressing null values, and dropping any necessary columns, we are ready to create our Random Forest Regression model.

From the sklearn package containing ensemble learning, we import the class Random Forest Regressor . The parameter *n\_estimators* creates *n* number of trees in your random forest, where *n* is the number you pass in. We passed in 10. The *.fit()* function allows us to train the model, adjusting weights according to the data values in order to achieve better accuracy. After training, our model is ready to make predictions, which is called by the *.predict()* method[5].

**Decision Tree Regression :** Decision Tree - Regression. Decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. Decision Tree is one of the most commonly used, practical approaches for supervised learning. It can be used to solve both Regression and Classification tasks with the latter being put more into practical application. It is a tree-structured classifier with three types of nodes.[6]

```

DecisionTreeRegressor

In [15]: from sklearn.tree import DecisionTreeRegressor
DT= DecisionTreeRegressor()
DT.fit(X_train,y_train)

predictD = DT.predict(X_test)

MAE= (mean_absolute_error(y_test,predictD))
print("MEAN ABSOLUTE ERROR VALUE IS :",MAE)
print(" ")

MSE=(mean_squared_error(y_test,predictD))
print("MEAN SQUARED ERROR VALUE IS :",MSE)
print(" ")

MedianAE=(median_absolute_error(y_test,predictD))
print("MEDIAN ABSOLUTE ERROR VALUE IS :",MedianAE)
print(" ")

EVS=(explained_variance_score(y_test,predictD)*100)
print("ACCURACY RESULT OF DECISION TREE REGRESSOR IS :",EVS)
print(" ")

R2=(r2_score(y_test,predictD))
print("R2 SCORE VALUE IS :",R2)
print(" ")

MEAN ABSOLUTE ERROR VALUE IS : 128593.037299138167
    
```

2(a)



```

print(" ")
EVS=(explained_variance_score(y_test,predictD)*100)
print('ACCURACY RESULT OF DECISION TREE REGRESSOR IS :',EVS)
print(" ")
R2=(r2_score(y_test,predictD))
print('R2_SCORE VALUE IS :',R2)
print(" ")
MEAN ABSOLUTE ERROR VALUE IS : 120583.03709130167
MEAN SQUARED ERROR VALUE IS : 46300104090.13283
MEDIAN ABSOLUTE ERROR VALUE IS : 65000.0
ACCURACY RESULT OF DECISION TREE REGRESSOR IS : 66.06391505761528
R2_SCORE VALUE IS : 0.6605934536942953

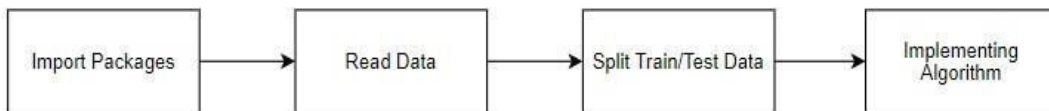
In [16]: import joblib
         joblib.dump(RF,'rf1.pkl')
Out[16]: ['rf1.pkl']

In [ ]:
    
```

2(b)

Fig 2(a) & 2(b), Performance measurements of Decision Tree – Regressor

**MODULE DIAGRAM:**



3(a)

Fig 3(a), Process to test/train data for implementing the algorithm

EXPERIMENTAL RESULTS :



4(a)

Fig 4(a), Prediction of house price web-page

#### IV. CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set is higher accuracy score is will be find out. This application can help to find the House Price. We have overseen out how to set up a model that gives clients for an original best methodology with look at future dwelling esteem expectations. Straight previous suggest works bring been used inside our model, something appreciate that that future worth expectations will have an inclination towards even more reasonable qualities. We composed an approach with use in basically the same manner as significantly data as time licenses for our expectation framework, by embracing those thoughts from guaranteeing inclination supporting. These consolidate redesigns we didn't choose as a result of compelled length of the time. A genuine concern for the expectation structure may be the stacking time frame. Additionally, our informational collection takes more than one day ought to get ready. As gone against playing out the calculations consecutively, we may use different processors and equal the calculations in question, which may potentially diminish the planning time Furthermore expectation period. Incorporate even more functionalities under the model, we can give decisions for customer with select a region on the other hand district should deliver rather than entering in the rundown.



## REFERENCES

- [1] John M. Clapp and Carmelo Giaccotto .” Evaluating House Price Forecasts” - 2002
- [2] Lynn Wu, Erik Brynjolfsson .” How Google Searches Foreshadow Housing Prices and Sales”- 2009
- [3] Vasilios Plakandaras, Rangan Gupta, Periklis Gogas and Theophilos Papadimitriou , “Forecasting the U.S. Real House Price Index” - 2013
- [4] Koen de Koning, Tatiana Filatova, Okmyung Bin,” Improved Methods for Predicting Property Prices in Hazard Prone Dynamic Markets” - 2016
- [5] Gaikwad Purva Chandrakant, Ganjave Pratiksha Namdev, Gorade Pooja Subhash, S. S. Gore, “Implementation of House Price Prediction Model Using Image Processing and Machine Learning” -2019



**SJIF: 8.423**



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INTERNATIONAL CENTRE



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details