



STUDY OF HOUSE PRICING PREDICTION USING PYTHON AND MACHINE LEARNING: IMPLEMENTATION

SK Althaf Hussain Basha ¹ A.Sucharitha ² N.V.Chandrika Naidu ³ P.Naga Mallika ⁴ K.Swetha ⁵
¹ Professor & Head, Krishna Chaitanya Institute of Technology & Sciences , Markapur, A.P, India
^{2,3,4,5} Scholar, Krishna Chaitanya Institute of Technology & Sciences , Markapur, India

ABSTRACT:

In this paper provides an overview about how to predict house costs utilizing different regression methods with the assistance of python libraries. The proposed technique considered the more refined aspects used for the calculation of house price and provided the more accurate prediction. It also provides a brief about various graphical and numerical techniques which will be required to predict the price of a house. In this paper contains what and how the house pricing model works with the help of machine learning and which dataset is used in our proposed model.

Keywords: Machine learning, Regression Technique, Classification Technique.

[1] INTRODUCTION

House/Home are a basic necessity for a person and their prices varying from location to location based on the facilities available like parking space, locality, etc. The house pricing is a point that worries a ton of residents whether rich or white collar class as one can never judge or gauge the valuing of a house based on area or offices accessible. Buying a house is one of the greatest and significant choices of a family as it expands the entirety of their investment funds and now and again covers them under loans. It is a difficult task to predict the accurate values of house pricing. Our proposed model would make it possible to predict the exact prices of houses.

In today's society, medical care problems have become a hot topic, and problems such as the unbalance and insufficient allocation of medical resources has become increasingly apparent. In this situation, the application of ML has become the unavoidable trend in the current development of medical care. As early as 1972, the scientists in the University of Leeds in the UK had been trying to use artificial intelligence (ANN) algorithms to judge abdominal pain. Now, more and more researchers are

committed to the combination of ML and medical care. The methods of pathological diagnosis of tumors, lung cancer, etc. by ML has gradually entered the field of vision. Some companies, such as Alibaba, Amazon, and Baidu have established their own research team working for it. This introduction of ML in medical care has greatly saved medical resources and provided a new way for citizens to see a doctor and facilitate people's lives. At the same time, the demand of people also provides a new important for the research and development of ML, with promoting its continuous improvement.

[2] LITERATURE SURVEY

Parul Kalra , Deepti Mehrotra et. al [1] studied for the social cause that was increasing at an alarming rate and was creating a situation of panic among the people of the world, Mortality Rate. This situation was analyzed by analyzing various factors such as birth rate, literacy rate, number of health centers, etc. using the decision tree technique in R tool which illustrated trees of two different decades separately and analyzed the factors affecting the mortality rate with their contribution in driving its rate, and also the summary of decision tree will indicate its accuracy and kappa factor to judge the authenticity of the factors chosen. This will be useful to the governing bodies to get to know about the factors and work upon them for the decrease of the infant mortality rate[2][3][6][7].

From both a local and international perspective, housing remains the most common form of land use. In recent times housing in many countries has rapidly evolved as an investment medium for private and institutional investors, which in turn is now demanding a higher level of research. The International Journal of Housing Markets and Analysis [16] aims to provide an international forum for the interchange of information and ideas relating to housing, housing markets[8][9][10].

Zhang et. al. [11] studied The objective of the empirically compare the predictive power of the hedonic model with an artificial neural network model on house price prediction. A sample of 200 houses in Christchurch, New Zealand is randomly selected from the Harcourt website. Factors including house size, house age, house type, number of bedrooms, number of bathrooms, number of garages, amenities around the house and geographical location are considered. Empirical results support the potential of artificial neural network on house price prediction, although previous studies have commented on its black box nature and achieved different conclusions[12][13][14][15].

Timor Kadir et. al. studied the Machine learning based lung cancer prediction models have been proposed [4] to assist clinicians in managing incidental or screen detected indeterminate pulmonary nodules. Such systems may be able to reduce variability in nodule classification, improve decision making and ultimately reduce the number of benign nodules that are needlessly followed or worked-up. In this article, we provide an overview of the main lung cancer prediction approaches proposed to date and highlight some of their relative strengths and weaknesses. We discuss some of the challenges in the development and validation of such techniques and outline the path to clinical adoption.

Coronal Mass Ejections (CMEs) are arguably the most violent eruptions in the Solar System. CMEs can cause severe disturbances in the interplanetary space and even affect human activities in many respects, causing damages to infrastructure and losses of revenue. Fast and accurate prediction of CME arrival time is then vital to minimize the disruption CMEs may cause when interacting with geospace. we build a prediction engine taking advantage of 182 previously observed geo-effective partial-/full-halo CMEs and using algorithms of the Support Vector Machine (SVM). We demonstrate that CAT-PUMA is accurate and fast. In particular, predictions after applying CAT-PUMA to a test set, that is unknown to the engine, show a mean absolute prediction error ~ 5.9 hours of the CME arrival time, with 54% of the predictions having absolute errors less than 5.9 hours. Comparison with other models reveals that CAT- PUMA has a more accurate prediction for 77% of the events investigated; and can be carried out very fast, i.e. within minutes after providing the necessary input parameters of a CME. A practical guide containing the CAT-PUMA engine and the source code of two examples are available in the Appendix, allowing the community to perform their own applications for prediction using CAT-PUMA[5]. Some authors and researchers are discussed and studied in House Pricing Prediction in similar areas of works

in Data Mining applications [17],[18],[19],[20],[21],[22] Machine Learning [23][24] , Big Data Analytics[25] , Text Classification[26], [27] and Deep Learning [28] domains areas for Prediction problems and Forecasting Problems[29].

[3] EXISTING SYSTEM

In The Existing system used xgboost for house price prediction. This study aims to explore the important explanatory features and determine an accurate mechanism to implement spatial prediction of housing prices in Beijing., based on the housing price and features data in Beijing, China. Our result shows that compared to traditional hedonic methods, machine learning methods demonstrate significant improvements on the accuracy of estimation despite that they are more time-costly. Moreover, it is found that XGBoost is the Less accurate model in explaining and predicting the spatial dynamics of housing prices in Beijing.

3.1 Disadvantages of Existing System

- In Xgboost, you have to manually create dummy variable/ label encoding for categorical features before feeding them into the models. Catboost/Lightgbm can do it on their own, you just need to define categorical features names or indexes.
- Training time is pretty high for larger datasets.
- Moreover, it is found that XGBoost is the Less accurate model in explaining and predicting the spatial dynamics of housing prices in Beijing.

[4] PROPOSED SYSTEM

The proposed method is based on the linear regression. This project is proposed to predict house prices and to get better and accurate results. The data for the house prediction is collected from the publicly available sources. In validation, training is performed on 50% of the dataset and the rest 50% is used for testing purposes. This technique splits the dataset into a number of subsets. At that point, it has been attempted for preparing on the entirety of the subsets; however, leave one (k-1) subset for the assessment of the prepared model. This strategy emphasizes k times with an alternate subset turned around for the preparation reason each time.

4.1 Advantages of Proposed System:

- The error free prediction provides better planning in the prediction of house price and other industries. This would be of great help for the people.
- This would be of great help to the people because the house pricing is a topic that concerns a lot of citizens whether rich or middle class as one can never judge or estimate the pricing of a house on the basis of locality or facilities available.
- Linear Regression is simple to implement and easier to interpret the output coefficients.
- The ability to determine the relative influence of one or more predictor variables to the criterion value.

4.2 System Architecture

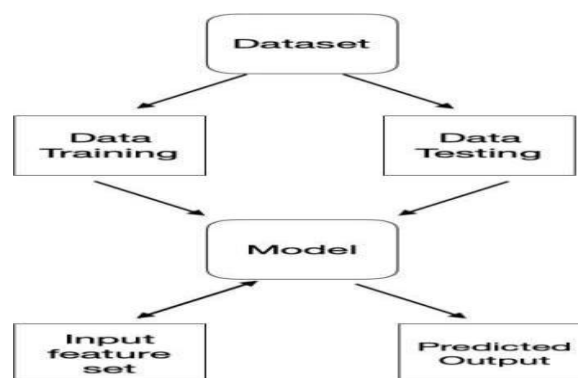


Fig. 1 Block Diagram for System Architecture

i) Site Analysis : Data collection is the process of gathering information on variables in a systematic manner. This helps in finding answers to many questions, hypothesis and evaluate outcomes. Data collection is the way toward social event and estimating data on focused factors in a built -up framework, which at that point empowers one to address pertinent inquiries and assess results. Information assortment is a part of research in all fields of study including physical and sociologies, humanities and business. While strategies differ by discipline, the accentuation on guaranteeing precise and legitimate assortment continues as before. It has been attempted for various datasets on Kaggle, which would suite our project objective. After looking at a lot of datasets, this dataset is found. It is a house pricing dataset in the city of Ames. This dataset is a very popular machine learning dataset with less scope of errors and variations.

ii) Data Visualization: House Pricing Prediction Using Python and Machine Learning. Data Visualization is the pictorial or graphical representation of information .It enables to grasp difficult concepts or identify new patterns. Data Visualization is seen by numerous orders as a cutting edge likeness visual correspondence. It includes the creation and investigation of the visual portrayal of information. To impart data plainly and effectively, information representation utilizes measurable illustrations, plots, data designs and different apparatuses. Effective visualization assists customers with separating and reason about data and verification. It makes complex data progressively accessible, reasonable and usable. Customers may have explicit logical endeavours, for instance, making assessments or getting causality, likewise, the structure standard of the reasonable (i.e., indicating examinations or demonstrating causality) follows the undertaking. Data Visualization is both a craftsmanship and a science. It is viewed as a piece of particular estimations by a couple, yet what's more as a grounded theory improvement device by others. Extended proportions of data made by Web activity and an expanding number of sensors in the earth are suggested as "enormous data" or Web of things. Dealing with, analyzing and passing on this data present good and orderly challenges for data portrayal. The field of data science and experts called data scientists help address this test.

iii) Data Pre-processing: It is the process of transforming data before feeding it into the algorithm. It is utilized to change over crude information into a clean dataset. It is an information mining strategy that includes moving crude information into a justifiable organization. The result of data pre-processing is the last dataset utilized for preparing and testing reason. Data pre- processing is an information mining procedure which is utilized to change the crude information in a helpful and productive format. In any Machine Learning procedure, Data Pre-processing is that progression wherein the information gets changed, or Encoded, to carry it to such an express, that now the machine can without much of a stretch parse it. Pre dealing with insinuates the progressions applied to our data before dealing with it to the estimation. Data Pre-processing is a system that is used to change over the rough data into an ideal enlightening assortment. In a manner of speaking, at whatever point the data is amassed from different sources it is assembled in rough setup which isn't feasible for the Examination. Genuine information for the most part contains clamours, missing qualities, and perhaps in an unusable organization which can't be legitimately utilized for Machine Learning models. Data pre-processing is required errands for cleaning the information and making it appropriate for an Machine Learning model which likewise expands the precision and proficiency of a Machine Learning model.

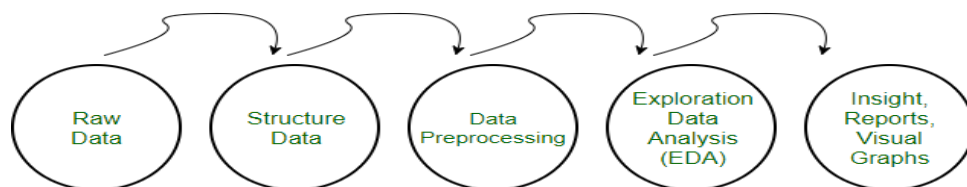


Fig.2 Data Pre-Processing

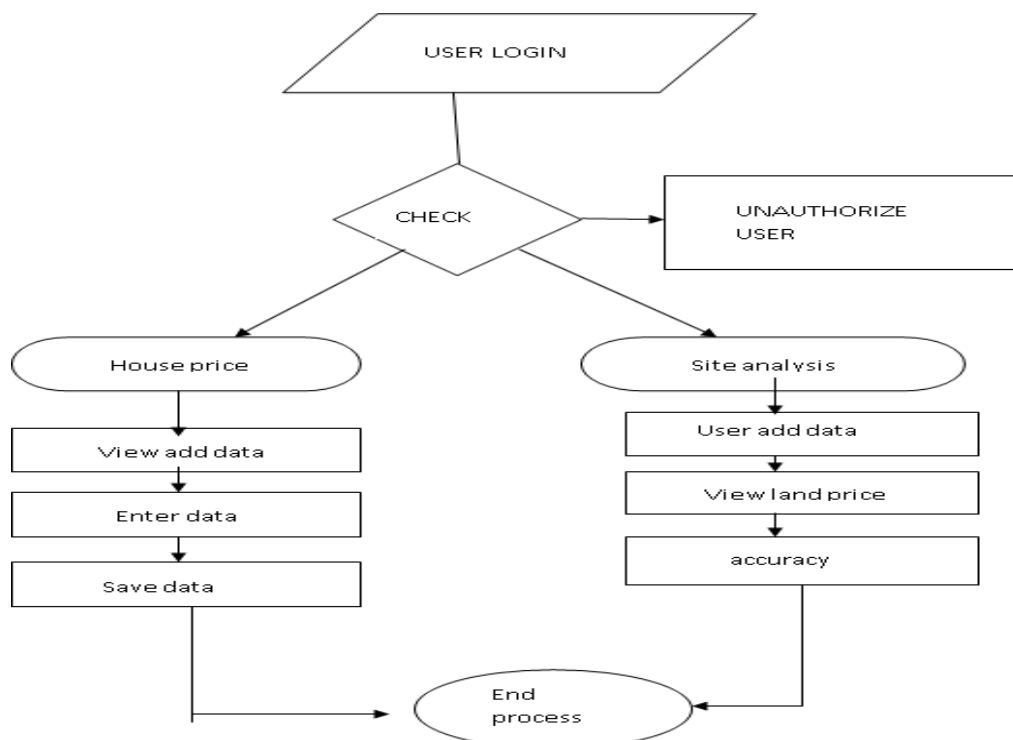


Fig. 3 User Login Process

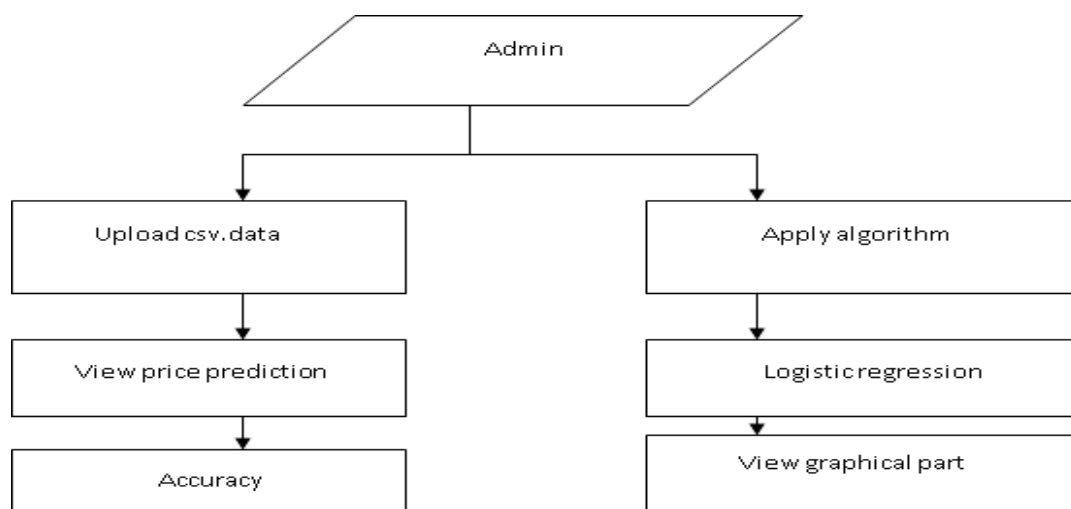


Fig. 4 Admin Process

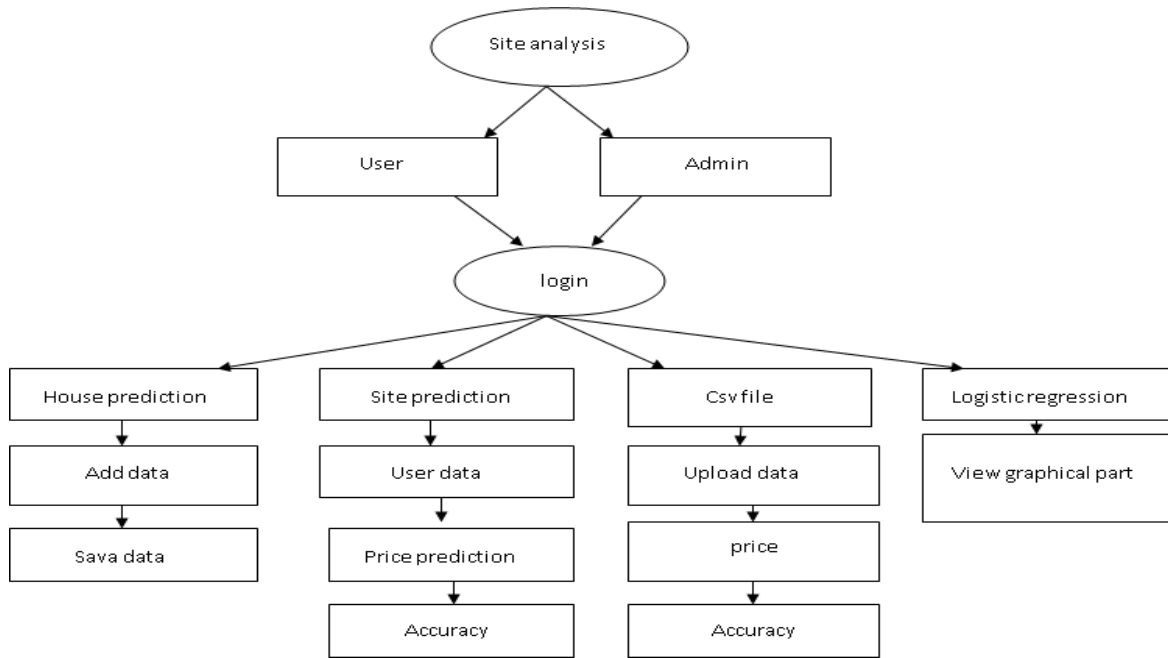


Fig.5 User and Admin Process

4.3 SOFTWARE ENVIRONMENT

In this paper we are implemented source code in Python Programming language and Django. Python is a general - purpose interpreted, interactive, object -oriented, and high-level programming language. An interpreted language, Python has a design philosophy that emphasise the code readability (not ably using whitespace indentation to delimit code blocks rat her than curly bracket s or keywords), and a syntax that allows programmers to express concept s in fewer lines of code than might be used in languages such as C++ or Java. It provides construct s that enable clear programming on both small and large scales. Python interpreters are available for many operating systems. CPython, the reference implement at i ion of Python, is open source soft ware and has a community- based development model, as do nearly all of its variant implementations. CPython is managed by the non- profit Python Soft ware Foundation. Python features a dynamic type system and automatic memory management . It support s multiple programming paradigms, including object - oriented, imperative, functional and procedural , and has a largeand comprehensive standard library. Django is a high-level Python Web framework that encourages rapid development andclean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source. Django's primary goal is to ease the creation of complex, database-driven websites. Django emphasizes reusability and "plug ability" of components, rapid development, and the principle of don't repeat yourself. Python is used throughout, even for settings files and data models.

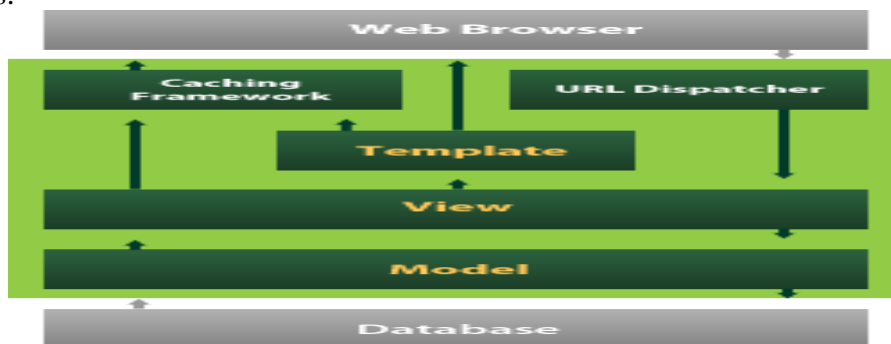


Fig 6. Design Pattern Used in Django-Model-View-Template-Controller Architectural

Pattern

Django also provides an optional administrative create, read, update and delete interface that is generated dynamically through introspection and configured via admin models

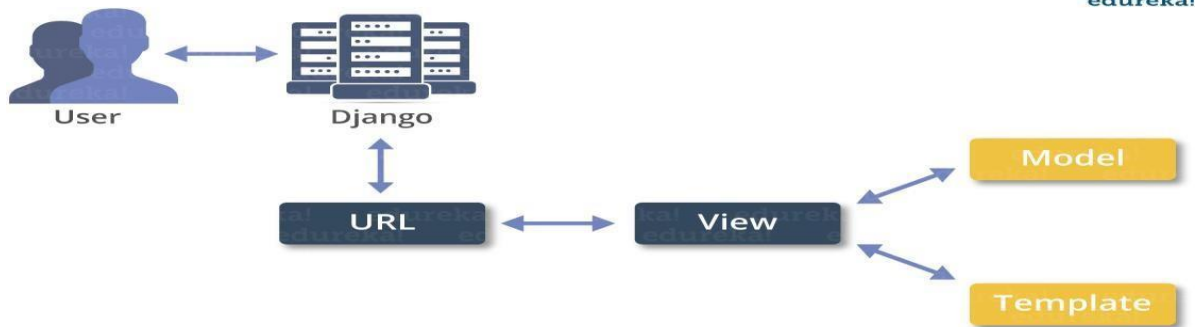


Fig. 7 :Model-View-Template(MVT) Architecture

[5] IMPLEMENTATION

5.1 Modules Description

i) User: The User can register the first. While registering he required a valid User email and mobile for further communications. Once the User registers, then the admin can activate the User. Once the admin activates the User then the User can login into our system. After login User will add the data to predict house values.

ii) Admin: Admin can login with his credentials. Once he logs in he can activate the users. The activated users only login in our applications. The admin will store csv data into our database. we can implement logistic algorithm to predict house and also we can perform cross validation.

iii) Machine learning: Machine learning refers to the computer's acquisition of a kind of ability to make predictive judgments and make the best decisions by analyzing and learning a large number of existing data. The representation algorithms include deep learning, artificial neural networks, decision trees, enhancement algorithms and so on. The key way for computers to acquire artificial intelligence is machine learning. Nowadays, machine learning plays an important role in various fields of artificial intelligence. Whether in aspects of internet search, biometric identification, auto driving, Mars robot, or in American presidential election, military decision assistants and so on, basically, as long as there is a need for data analysis, machine learning can be used to play a role.

5.2 Screen Shots

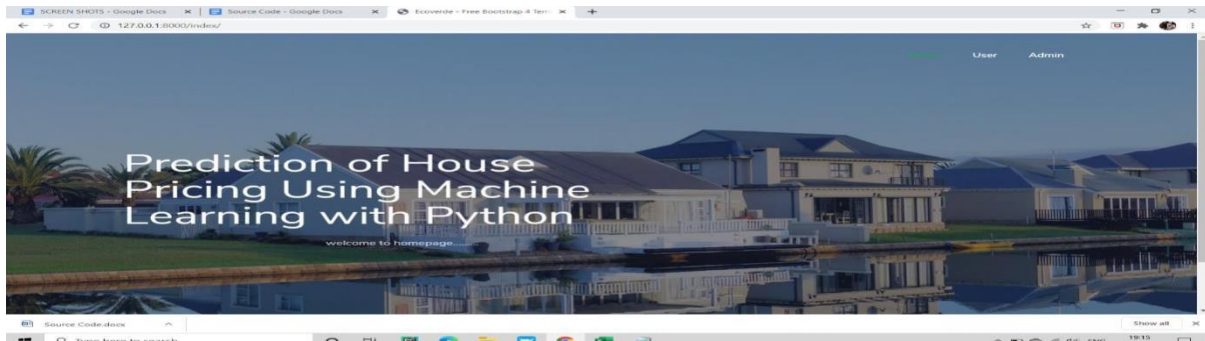


Fig. 8 Home Page

In the above screen, home page is displayed.

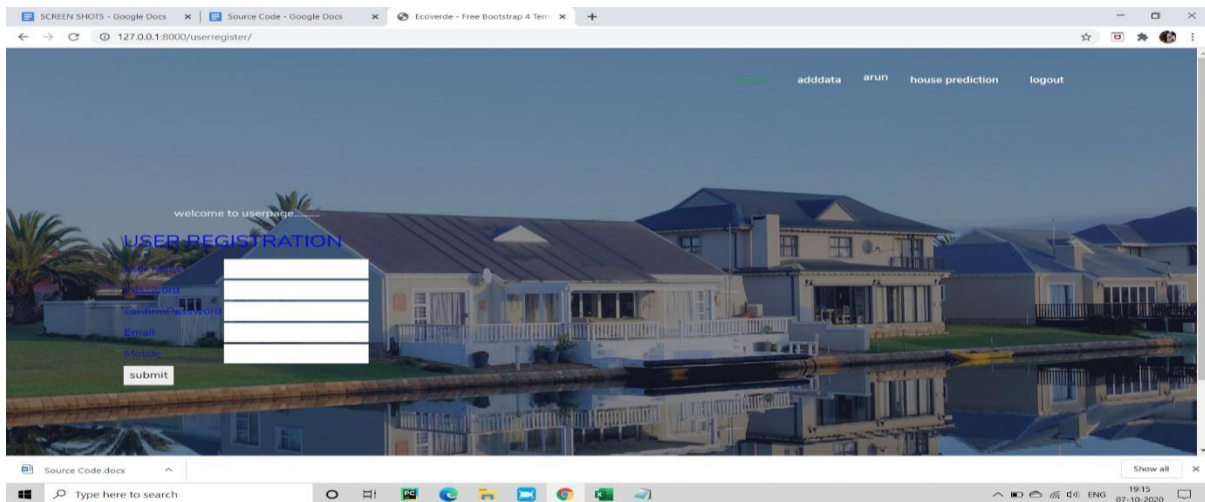


Fig. 9 User Register Page

In the above screen , we see the user registration form which contains username, password , confirm password, e-mail , mobile.

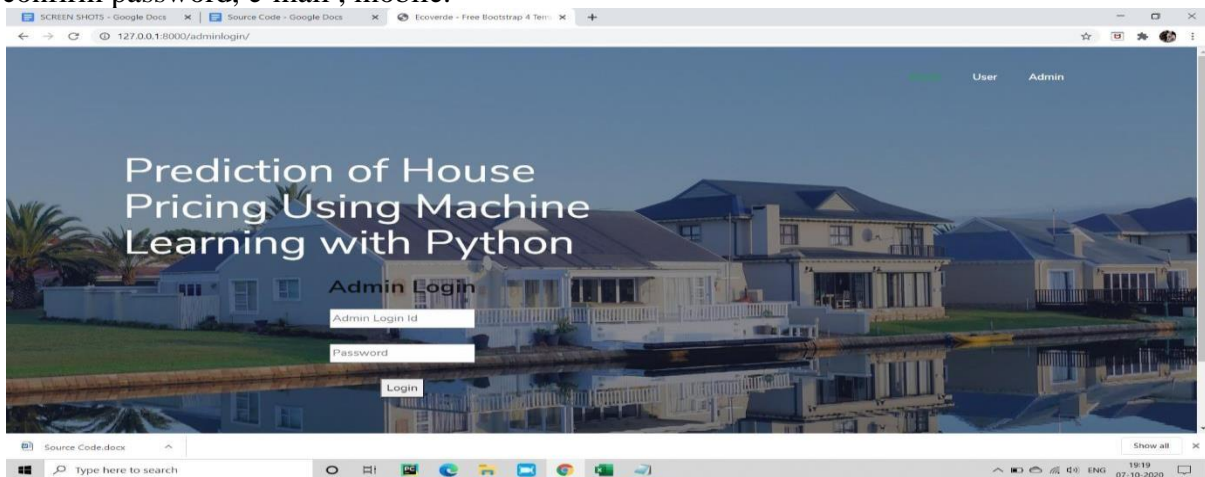


Fig.10 Admin Login Page

In the above screen , we see the admin login page which contains admin login id and password.

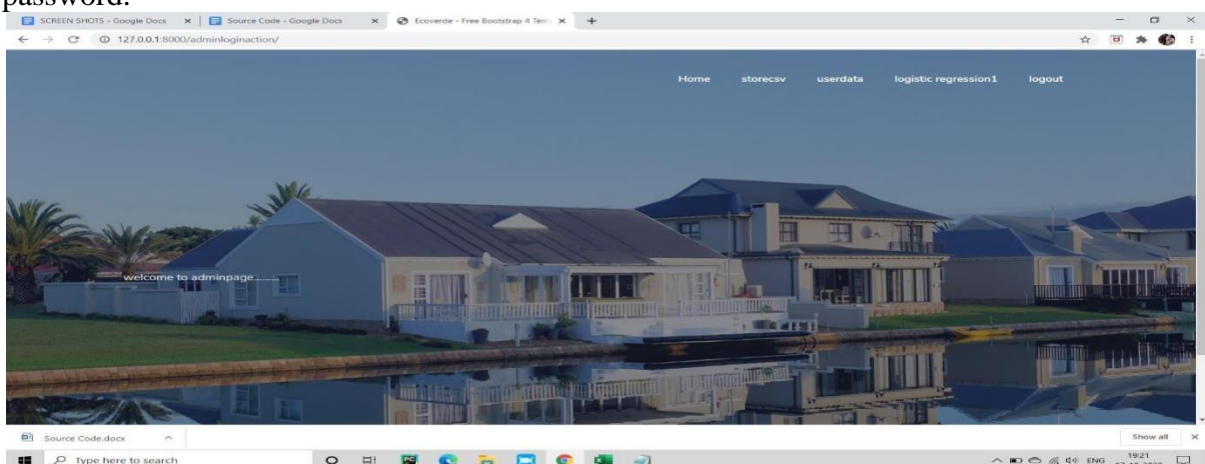


Fig. 11 Admin Home page

After admin login, admin home page is displayed which contains home, store csv , user data , Logistic regression and logout.

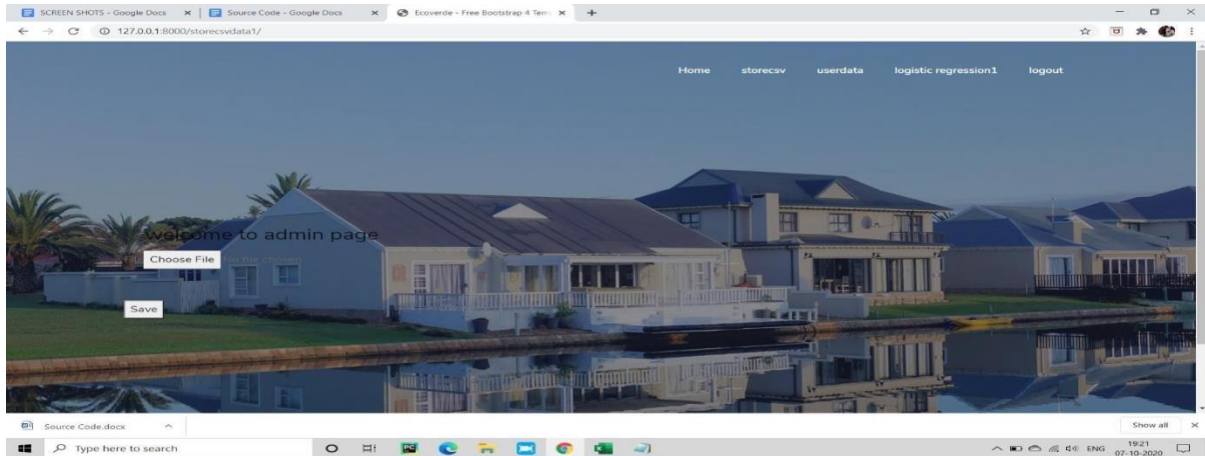


Fig. 12 Store /Upload csv format file

In the above screen, we can store a csv Data.It means a Excel file can store as a csv data

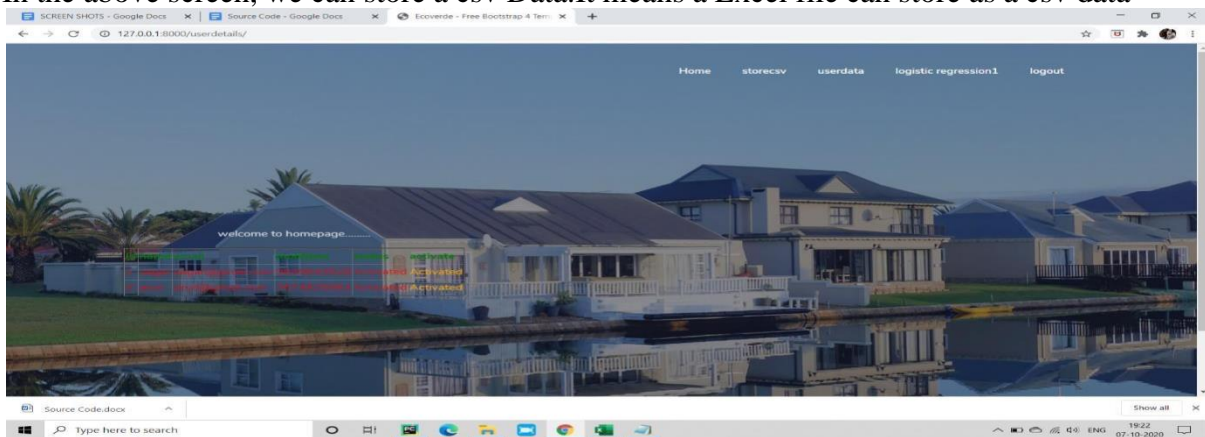


Fig. 13 User Data Page

In the above screen , Once the user registers , then the admin can activate the user .once the admin activates the user then the user can login into our system.

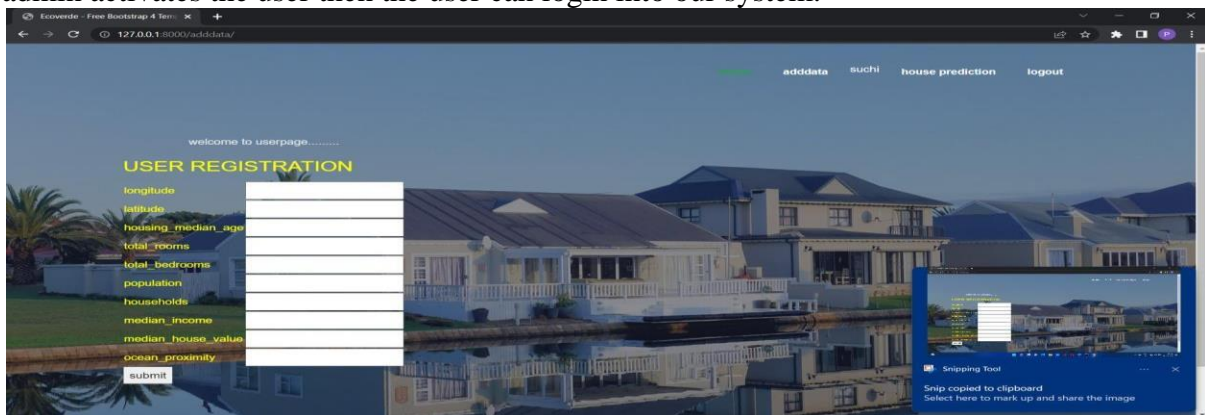


Fig. 14 Add Data Page

In the above screen, we can add values to the particular data.

```
C:\Windows\System32\cmd.exe - python manage.py runserver
0 longitude 20640 non-null float64
1 latitude 20640 non-null float64
2 housing_median_age 20640 non-null float64
3 total_rooms 20640 non-null float64
4 total_bedrooms 20640 non-null float64
5 population 20640 non-null float64
6 households 20640 non-null float64
7 median_income 20640 non-null float64
8 <1H OCEAN 20640 non-null float64
9 INLAND 20640 non-null float64
10 NEAR BAY 20640 non-null float64
11 NEAR OCEAN 20640 non-null float64
dtypes: float64(12)
memory usage: 1.9 MB
predictions: [238609.3213121 381306.38850169 257314.61040085 131301.84992786
185894.87077418 539843.34729246 125404.98781123 441654.60911985
160118.94728117 201143.53254885]
error diff:
predicted Actual Diff
0 238609.321312 235000.0 3609.321312
1 381306.388502 386800.0 -5493.611498
2 257314.610401 286400.0 -29085.389599
3 131301.849928 120100.0 11201.849928
4 185894.870774 114200.0 71694.870774
5 539843.347292 500001.0 39842.347292
6 125404.987811 68700.0 56704.987811
7 441654.609120 500001.0 -58346.390880
8 160118.947281 134100.0 26018.947281
9 201143.532549 143000.0 58143.532549
[03/Jun/2022 12:08:14] "GET /houseprediction/ HTTP/1.1" 200 39288
```

Fig. 15 Output / Result Page

[6] CONCLUSION & FUTURE WORK

The sales price for the houses are calculated using different algorithms. The sales prices have been calculated with better accuracy and precision. This would be of great help for the people. To achieve these results, various data mining techniques are utilized in python language. The various factors which affect the house pricing should be considered and work upon them. Machine learning has assisted to complete out task. Firstly, the data collection is performed. Then data cleaning is carried out to remove all the errors from the data and make it clean. Then the data preprocessing is done. Then with help of data visualization, different plots are created. This has depicted the distribution of data in different forms. Further, the preparation and testing of the model are performed. It has been found that some of the classification algorithms were applied on our dataset while some were not. So, those algorithms which were not being applied on our house pricing dataset are dropped and tried to improve the accuracy and precision of those algorithms which were being applied on our house pricing dataset. To improve the accuracy of our classification algorithms, a separate stacking algorithm is proposed. It is extremely important to improve the accuracy and precision of the algorithms in order to achieve better results. If the results are not accurate then they would be of no help to the people in predicting the sales prices of houses. It also made use of data visualization to achieve better accuracy and results. The sales price is calculated for the houses using different algorithms. The sales prices have been calculated with better accuracy and precision. This would be of great help for the people.

In future, many more algorithms can be applied on this dataset such as decision tree, Naïve Bayes, SVM etc. and find out their respective accuracies and use them to predict a better outcome and hence increase the accuracy. The KNN algorithm can also be applied to predict the accuracy. The k-means algorithm can also be applied. With the help of these algorithms, the house prices are accurately predicted. Hence, it would be of great help for the government and the people themselves. Regression algorithms are initially took up for our project but in the future, this can also be achieved using the classification algorithms. The classification algorithms can be used and it can also be applied to our house pricing dataset and see if they are being applied properly or not. The accuracy and precision of these algorithms can also be improved according to our needs. They can choose the house that best suits their budgets so that they don't have to take any kind of loan from the banks. In the future, an application can also be developed for the same. That would make it even easier for the people to select them houses that best suits their budgets.

Artificial intelligence can also be applied to make our project more enhanced in the future. More factors that can affect the house pricing of a particular area will also be considered. House pricing of an area can also depend upon political and emotional factors. It can be affected by the direction in which the house points. People believe that facing of a house is extremely important. Some believe that the house number should make a total of either 3 or 8. All these factors are taken into consideration in the future and work upon them to mark our project stronger and more relevant for public use. although many algorithms are used in our project still, many more regression and Page classification algorithms are used to top make our project and make it more helpful for the people. Various methodologies from the field of machine learning are used to make our project more relevant. Sometimes people also prefer to stay near areas where basic facilities are easily available such as a general store, mother dairy, photocopy shop etc. This is also an extremely important factor that may affect the prices of houses and can be taken into consideration in the future. All the major factors that can affect the prices of houses in a particular area are almost covered and have worked upon them. In the future, some of the minor factors that can affect house pricing on a smaller scale can be identified and can work upon them that how do they affect house pricing and what can be done to minimize it. In the future the model deployment of more algorithms can be performed to achieve accurate results.

REFERENCES

- [1] Jain, N., Kalra, P., & Mehrotra, D. (2019). Analysis of Factors Affecting Infant Mortality Rate Using Decision Tree in R Language. In *Soft Computing: Theories and Applications* (pp. 639-646). Springer, Singapore.
- [2] R. A. Rahadi, S. K. Wiryono, D. P. Koesrindartotoor, and I.B. Syamwil, —Factors influencing the price of housing in Indonesia,|| *Int. J. Hous. Mark. Anal.*, vol. 8, no. 2, pp. 169–188, 2015
- [3] V. Limsombunchai, —House price prediction: Hedonic price model vs. artificial neural network,|| *Am. J.*, 2004
- [4] Kadir, T., & Gleeson, F. (2018). Lung cancer prediction using machine learning and advanced imaging techniques. *Translational Lung Cancer Research*, 7(3), 304-312.
- [5] Liu, J., Ye, Y., Shen, C., Wang, Y., & Erdélyi, R. (2018). A New Tool for CME Arrival Time Prediction using Machine Learning Algorithms: CATPUMA. *The Astrophysical Journal*, 855(2), 109.
- [6] Velankar, S., Valecha, S., & Maji, S. (2018, February). Bitcoin price prediction using machine learning. In *Advanced Communication Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020)* IEEE Xplore Part Number: CFP20V66-ART; ISBN: 978-1-7281-4108-4 978-1-7281-4108-4/20/\$31.00 ©2020
- [7] Malhotra, R., & Sharma, A. (2018). Analyzing Machine Learning Techniques for Fault Prediction Using Web Applications. *Journal of Information Processing Systems*, 14(3).
- [8] Choo, M. S., Uhm, S., Kim, J. K., Han, J. H., Kim, D. H., Kim, J., & Lee, S. H. (2018). A Prediction Model Using Machine Learning Algorithm for Assessing Stone-Free Status after Single Session Shock Wave Lithotripsy to Treat Ureteral Stones. *The Journal of urology*.
- [9] Nilashi, M., Ibrahim, O., Ahmadi, H., Shahmoradi, L., & Farahmand, M. (2018). A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. *Biocybernetics and Biomedical Engineering*, 38(1), 1-15.
- [10] Fan, C., Cui, Z., & Zhong, X. (2018, February). House Prices Prediction with Machine Learning Algorithms. In *Proceedings of the 2018 10th International Conference on Machine Learning and Computing* (pp. 6-10). ACM.
- [11] Zhou, J., Zhang, H., Gu, Y., & Pantelous, A. A. (2018). Affordable levels of house prices using fuzzy linear regression analysis: the case of Shanghai. *Soft Computing*, 1-12.
- [12] Jang, H., Ahn, K., Kim, D., & Song, Y. (2018, June). Detection and Prediction of House Price Bubbles: Evidence from a New City. In *International Conference on Computational Science* (pp. 782-795). Springer, Cham.
- [13] Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7), 1145- 1159.
- [14] Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert Systems with Applications*, 42(6), 2928-2934.
- [15] Harrison, D., and D. L. Rubinfeld. 1978. "Hedonic Housing Prices and the Demand for Clean Air." *J. Environ. Econ. Manag.* 5 (1): 81– 102.
- [16] RJ Bolton, DJ Hand , House Prices and Analysis ,*International Journal of Housing Markets and Analysis*.
- [17] Sk Althaf Hussain Basha, A.Govardhan ,Y.R.Ramesh Kumar , Mohd. Zaheer Ahmed, "Predicting Student Academic Performance Using Temporal Association Mining" *International Journal of Information Systems and Education (IJISE)*, Vol.2, No.1(2012),pp.21-41,ISSN: 2231- 1262.

- [18] Sk. Althaf Hussain Basha, A.Govardhan “Rank Analysis Through Polyanalyst using Linear Regression” , International Journal of Computer Science and Network Security(IJCSNS), VOL.9 No.9,2009, pp. 290-293, ISSN: 1738-7906.
- [19] GNR Prasad, SK Althaf Hussain Basha, , Mallikharjuna Rao K M GnanaVardhan “A Review of Predictive And Descriptive Data Mining Techniques In Higher Education Domain, International Journal of Computer Engineering and Applications(IJCEA),Volume 13, Issue 6, January. 21, ISSN2321-3469
- [20] Y. Sri Lalitha, Sk. Althaf Hussain Basha, Y. Gayatri and M. V Aditya Nag, "Data Driven Prediction Model for Crime Patterns", Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough, Studies in Computational Intelligence 956, Springer.
- [21] Y Ramesh Kumar, Sk Althaf Hussain Basha, Y Vijayalata, “ Predicting Student Academic Performance using Temporal Association Mining-A case Study on Educational Data ” , IEEE International Conference on Engineering for Humanity (ICEH-2013), Hyderabad, pp. 21- 27,2013, ISSN:978-93-82880-53-0.
- [22] Donapati Srikanth SK Althaf Hussain Basha, T Naveen Kumar , V. Anand , “Categorization of Academic Student Performance using Hybrid Techniques” International Conference on Advanced Computing Methodologies (ICACM-2013), Hyderabad, pp.325- 330,2013.
- [23] Ch. Prakash, Sk Althaf Hussain Basha, , D. Mounika, G. Maheetha, “An Approach for Multi Instance Clustering of Student Academic Performance in Education Domain”, IJJDWM Journal, Volume 3,Issue 1,pp.1-9, Feb.2013,ISSN: 2249-7161
- [24] SK Althaf Hussain Basha, A.Govardhan, “MICR: Multiple Instance Cluster Regression for Student Academic Performance in Higher Education”, International Journal of Computer Applications(IJCA), Volume 14– No.4,2011,pp.23-29, ISSN: 0975-8887
- [25] Ayesha Mariyam, Sk. Althaf Hussain Basha, "An Approach To Big Data Analytics And Its Significance", Online International Interdisciplinary Research Journal(OIIRJ), Volume-07, May 2017 Special Issue, pp.75-83, ISSN2249-9598.
- [26] Ayesha Mariyam, SK Althaf Hussain Basha, and S Vishwanadha Raju "Applications of Multi- Label Classification", International Journal of Innovative Technology and Exploring Engineering(IJITEE),pp.86-89,ISSN:2278-3075, Volume-9 Issue-4S2, March 2020,Blue Eyes Intelligence Engineering & Sciences Publication.
- [27] P M Yohan, Sk Althaf Hussain Basha, B Sasidhar, A.Govardhan , “ Automatic Named Entity Identification and Classification using Heuristic Based Approach for Telugu”, IJCSI(International Journal of Computer Science Issues), Volume 10, Issue 6,November 2013,ISSN: 1694-0814
- [28] Ayesha Mariyam, SK Althaf Hussain Basha, S Viswanadha Raju, Y Sri Lalitha, "A Brief Research On Deep Learning Models", International Journal of Computer Engineering and Applications, Volume 13, Issue 6, November 2020, ISSN2321-3469.
- [29] B Sasidhar, Sk Althaf Hussain Basha , “ A Comparative Study of Educational Data Mining Methods Used to Forecast Student Success and Failures”, International Journal Computer Science Information and Engineering Technologies (IJCSIET), International Conference 2014,ISSN:2277-4408,2014
- [30] Dr. G. N. R. PRASAD, “Identification of Bloom’s Taxonomy level for the given Question paper using NLP Tokenization technique”, Turkish Journal of Computer and Mathematics Education, Vol.12 No.13 (2021), 1872-1875.