International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. **05** Issue **01** | **2020**

Price Estimator Using Data Science & Machine Learning

Mr. Maheshwar M
MCA Scholar, School of CS & IT, Dept of MCA,Jain (Deemed-to-be) University, Bangalore

ABSTRACT

Housing Price patterns are the worry of purchasers and dealers; however, it additionally demonstrates the current financial circumstance. There are numerous elements that affect house costs, for example, the number of rooms and restrooms. Indeed, even the close by area, an area with an extraordinary openness to roadways, interstates, schools, shopping centers, and neighborhood business openings adds to the ascent in house costs. Manual house forecast gets troublesome, thus there are numerous frameworks created for house value expectation. As we live in the Bangalore Metropolitan city it is a very tedious cycle to locate the home in the necessary zone and locate the specific estimation of the house dependent on the area. The model which I have proposed utilizes Linear Regression. This framework point is to cause a model which to can give us a decent house price expectation dependent on various elements like region, rooms, restrooms, and so forth. The dataset utilized in this models contains the subtleties of houses in Bangalore areas with different variables, where we will play out the different Data science ideas like Data Cleaning to Remove the unimportant information, locate the invalid qualities and handle the invalid qualities, finding the exceptions, and taking care of it, and we will Build the Module with the Simple Linear Regression model using python and This Price Estimator Project has two modules, in particular, Admin and Normal User. The Admin has Complete access to the Web Application & Database, he can manage the lists of users. The Normal Users only use the Estimator Functionality & update the Password. The Objective of the Project is to help real-estate clients who will in general purchase the house, comes very handy and save the time, let us know the price of the house just with few taps.

Keywords: - Machine Learning, Dataset, Pandas, Linear Regression, Data Cleaning, Outlier Detection, SQLite3, Admin User, Normal User, Python Flask Web-Framework, Sklearn, Prediction System.

1. INTRODUCTION

Data is the core of the Machine Learning. Predictive models use Data for Training which gives fairly precise outcomes. Without Data, we can't prepare the model. Machine learning includes building these models from Data and utilizations them to foresee new information. Machine Learning is a subset of Artificial Intelligence. It gives framework capacity to learn wherein it consequently learns and improves its Performance without being expertly programmed. It centers around the advancement of projects and use it to learn for themselves. As the world is pushing ahead to utilizing variations advancements, so has automation improved its approaches to make our work simpler. Machine Learning is firmly identified with Statistics, which centers around making expectations utilizing PCs. There are various varieties of applications for Machine Learning, for example, email filtering, where it is hard to build up an ordinary calculation to play out tasks efficiently. Machine Learning Algorithms are purely based on the data. AI calculations are a serious variant of the regular algorithms. It makes programs "more astute" by permitting them to consequently gain from the information given by us. The algorithm is predominantly separated into two stages and that is the Training stage and the testing stage. Extensively three kinds of Algorithms are essentially utilized on information and they are supervised, unsupervised and reinforcement learning algorithms.In Supervised learning, the algorithms comprised of an objective(target) variable or a dependent variable which is to be anticipated from a lot of independent Variables. Utilizing a Function, the inputs are planned to the ideal yields. Instances of Supervised learning algorithms are Decision Tree, Regression, Random Forest, Logistic Regression, KNN and so on. In Unsupervised learning, the calculation doesn't have any dependent variables. It is utilized for bunching into various groups. K-Means, Independent Component Analysis, Apriori Algorithms, Principal Component Analysis are a few instances of Unsupervised learning calculations. Reinforcement Learning is utilized when the machine is utilized to settle on explicit choices. In this, the model is in an environment where it trains itself making it more exact by utilizing the experimentation technique. The model subsequently gains from past encounters and it catches the information about that domain to settle on accurate decisions, Example of Reinforcement Learning: Markov Decision Process-hot encoding is one such Reinforcement learning Algorithm.Regression Analysis is a type of prescient displaying method which researches the connection among needy and autonomous variable. Regression Analysis includes diagramming a line on a bunch of data points that most intently fits the overall state of the information or regression shows the adjustments in a Dependent variable on the y-axis to the adjustments in the Independent variable on the x-axis. Regression can be utilized to determine the quality of Predictors, Trend Forecasting, and Forecasting Effects.

There are Different Types of Regression like Linear Regression, Polynomial Regression, Logistic Regression, Lasso Regression, Decision Tree Regression, Random Forest Regression. I have Implement the Linear Regression in the Project. Linear Regression is an approach to modeling the relationship between dependent variable and one or more independent variable by fitting a linear equation to observed data.

2. LITERATURE REVIEW

[1] The Author in this paper proposed a technique on how we can break down and foresee the house costs utilizing the different Machine Learning models and permits the clients to select on their own decisions dependent on their necessities and select a model which coordinates their prerequisites. [2] The Author in this paper predicts presidential housing costs utilizing three AI strategies Regression, Random-Forest & support vector machine. In this paper, we can see that the author analyzes some miniature elements, similar to the Pool area, Lot area, that can be utilized as highlights to anticipate house cost. We fit a somewhat basic Simple Regression Model that contains a couple of qualities of a private resource, and we can arrive at a genuinely decent outcome. All the more explicitly, microeconomics points of view are presented in this model. Highlights like Pool area, Lot area, fundamentally influences the valuing of private resource evaluating.

[3] This Author in this article for the most part focuses on the examination between various AI calculations (Angle boosting, LASSO, Ada Boosting Regression, Multiple Linear Regression, Regression, Ridge Regression, Elastic Net Regression) about House value expectation Analysis. From the above investigation results, the gradient boosting algorithm has high exactness esteem when contrasted with the wide range of various calculations in regards to house value expectations. [4] In this paper, a few tests have been performed utilizing linear regression and particle swarm optimization methods to perform house value expectations. In light of the NJOP information of 9 houses, the framework is demonstrating house value expectations into 7 models every one of them speaks to one area. Out of these 7 Models, just one model with the zone Karang Besuki region gets the best boundary for the ideal forecast, with just 0.8 exactness. For the other model, the mistake forecast esteems are still huge. Utilizing various strategies that coordinate the time-arrangement information will be utilized in future examination to get more modest mistake forecast esteems and utilizing more information to improve results

3. PROBLEM STATEMENT

House prices are a significant impression of the economy, and Housing value ranges are of Great Interest for both then Vendors & buyers. In this project, house costs will be anticipated based on the given features (Dependent Variables) that spread numerous aspects of residential houses. The dataset is the costs and highlights of Residential houses in Bangalore, India, acquired from the Kaggle Competition. This dataset comprises of Basic house Features and 13320 houses with sold costs. It contains some highlighting Features, for example, the Area of the houses, Number of rooms, and Number of restrooms. Such highlights empower us to investigate different strategies to foresee house costs. In this Project we are using the Simple Linear Regression Analysis, To the Determine the Relationship of the variable price with the Four predictor variables Total Square feet of the Houses, Number of Bathrooms, Number of Bedrooms (BHK), Locations. Hear we need to determine the how the different values of the predictor variable affect the value of the response variable price.

4. SOLUTION

The proposed method Uses the concepts of the Data Science concepts like Data Cleaning, Outlier Detection & Linear Regression to analyze the relationship between the Variable and Estimate the Price accordingly & It uses a Python Flask Web Framework to create a Web-Based Application.

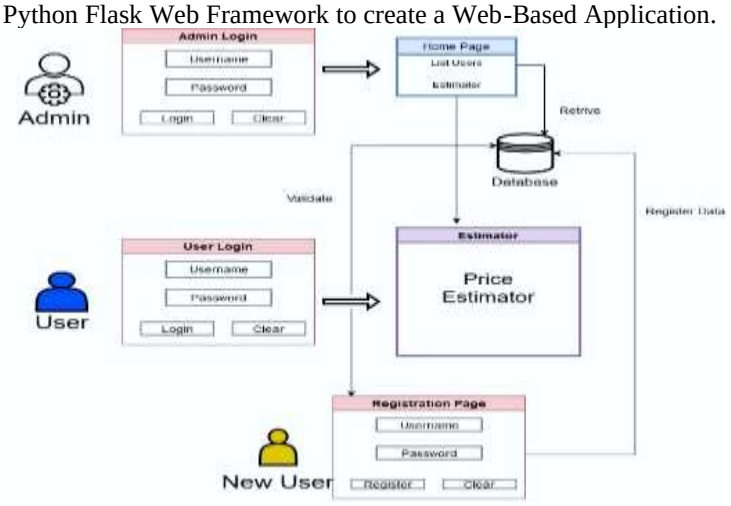

Fig 4.1 System model

The Flowchart represents the work flow of the Estimator Application

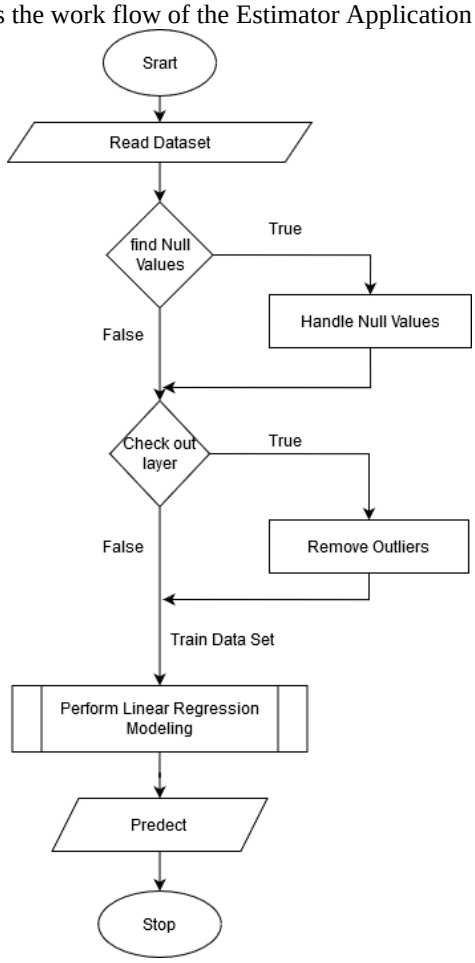

Fig. 4.2 Flowchart

4.1 Features

- Dataset Collection Process: I have Collected the Dataset by enrolling in the Kaggle Bangalore House Price Prediction competition.
- Dataset Cleaning Processing: We have used the Pandas library to Clean the data and make it ready for processing. Removing the variables which does not have linear relation between the response variable, processing the null values, splitting the variables.
- Dataset Modelling: I have used the Simple Linear Regression to build the build the Model. Once the Model is build, we can store all the data in the pickle file which can be use by the Python Web Frame Work for Processing.
- Web Framework & Database: I have used the Flask Server to host the Web application in the local PC.
- Web Application Architecture:
  a. Home Page:
    i. Login Page
    ii. Registration Page
    iii. Admin Login
- Login Page: Retrieves the data form the Database (SQLite3) and validated the user if the user is not validated then it will throw the error.
- Registration Page: It collects the basic details & password for registration.
- Admin Page: After Validating the User, it provides functionalities like managing the user and using the main app.

4.2 Software required

- Python 3.6
- OpenCV 4.1
- SQL lite3

4.3 Modules

- Sklearn
- Flask
- SQLite
- Pickle
- Json
- Pandas
- Matplotlib

5. EXPERIMENT

- It uses python's panda's library to perform the data cleaning operations like removing Null values, removing outliers.
- It uses Linear Regression Model to determine the price of the Houses.
- After the model is been created successfully it is stored in a pickle file to which stores the values in the binary format
- Later the pickle files are used in backend server as dataset to perform the calculations.
- It uses SQLite3 database to store the user credentials which will be used to login through console.
- There are 2 types of users Admin and Normal user
- Admin has the privileges to access all the Functionalities of the Application like Creating the User details, updating user details, deleting user details & access to the Estimator Page.
- Normal User Once the uses is logged in, he has the assess to the Estimator Page where he can predict the price of the residential houses in Bangalore.
- The application is hosted using a python's Flask Web Framework which is the Development Environment for the Web Application.

6. RESULT ANALYSIS

- Login page which gathers the information from the User and Validates the User, in the event that the client is approved, at that point it will explore the client to the Main Application (Estimator). In the event that the Username and the Password aren't right, at that point, it tosses a message username or password not accessible.

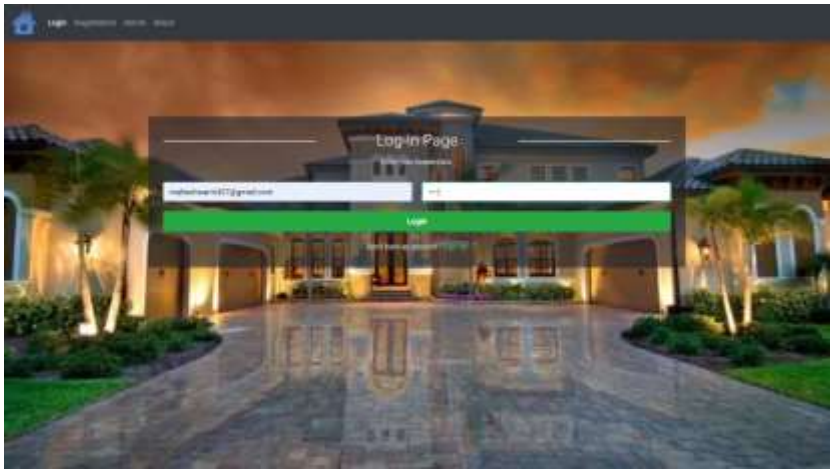

Fig 6.1 Login Page

- This is the Main Application that predicts the cost of the houses in Bangalore Location. All the area subtleties are been taken from the .json file and the Model is built using Linear Regression and executed utilizing Python and put away in the pickle file and this application utilizes that pickle file to gauge the cost of the houses.

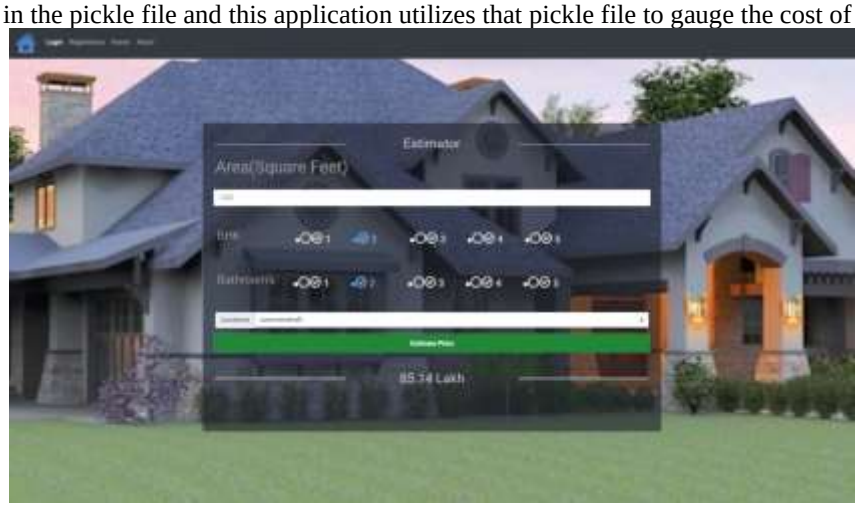

Fig 6.2 Estimator (Main Application)

- Registration Page gathers the information from the client and stores the information in the Database.

Fig 6.3 Registration Page

- Administrator landing page permits the administrator to explore through the application and deal with the clients.

Fig 6.4 Admin Home page

- View user page retrieve all the details of the users for the database and list them only admin can access this page.

Fig 6.5 List Users

- About page contains brief description of the Project and the Developer.

Fig 6.6 About Page

7. CONCLUSION & FUTURE ENHANCEMENT

This paper observationally thinks about the prescient intensity of the model form utilizing the Linear Regression which has an exactness of 84.52% which is best among the Algorithms like Lasso which has a 68.4% and Decision tree with 72.31%.

Our model has a good Sklearn score, yet there is still opportunity to get better. In a certifiable situation, we can utilize such a model to anticipate house costs. This model should check for new information, once in a month, and consolidate them to extend the dataset and produce better outcomes.

We can evaluate other progressed relapse methods, similar to Random Forest and Bayesian Ridge Algorithm, for Prediction's. Since the information is exceptionally connected, we ought to likewise attempt Elastic Net relapse procedure.

8. REFERENCES

[1] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
[2] https://machinelearningmastery.com/linear-regression-for-machine-learning/
[3] G. Naga Satish, Ch. V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu, "House Price Prediction Using Machine Learning, 2019"
[4] Yiyang Luo, "Residential Asset Pricing Prediction using Machine Learning, 2019"
[5] Anuradha G, CH.Raga Madhuri, M.Vani Pujitha, "House Price Prediction Using Regression Techniques: A Comparative Study, 2019"
[6] Hilman Taufiq, Wayan Firdaus Mahmudy, Adyan Nur Alfiyatin, Ruth Ema Febrita, "Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization, 2017"