

Artificial Intelligence- and Machine Learning-Based House Price Prediction Using IoT HetNets

¹Nishant Sharma, ²Mohd Safeen khan
^{1,2} Galgotias University, Greater Noida

Submitted: 25-05-2021

Revised: 01-06-2021

Accepted: 05-06-2021

ABSTRACT: In today's world, machine learning plays a very important role in various fields to make user interaction and experience better and feasible in image detection, spam reorganization, and product recommendation and also in enhancing security alerts

The main purpose of this paper is to predict the economical house valuation of the properties of the customers with relation to their budgets and priorities. This will be a support for the people who are trying to purchase a house with the result from/of an establishment of the budgets of marketing them. Further for the prediction, the users require the current system which involves the computation of house costs while not necessary to predict for future market estate theme. This business totally depends upon a period of time, accuracy, and methods with a large amount of analytical information to create critical choices. With this potency and capability of a business, it will be easier to accomplish machine learning. The exchangeable intelligence, chat-bots, and automation helps to smooth out the business methods.

While surfing through social networking sites or other apps and services, you must have observed that the chat boxes pop up. Which in this case will be the bots, pre-trained to help customers instantly without including human intervention as far as possible. And probably give much better suggestions comparatively. In general house costs preserve increasingly terribly which proves that the house costs and unit typically exaggerate. On the other hand, this model can further be implemented to various Internet-based platform or smartphones in the form of application to make it user-friendly and more feasible with different functionalities using IoT HetNet.

Keywords Realtime, House price prediction, Data-set

I. INTRODUCTION

Artificial intelligence [1] plays an important role in building these models to form information and uses them to predict new information. As the users know Machine learning[2] is basically a huge set of data trained using complex mathematical algorithms to achieve desired results. Thus it provides a system to detect, spam reorganization without being explicitly

programmed. It also pays attention to the building method of programs to use it later for users, because the world is advancing itself the exploitation of numerous new technologies [3]. Therefore, automation enhances our ways which create our work to do it in an easy way. Although the word automation has been discovered within the 1950, only a few individuals understand its meaning. Robotic method of automation recommends which build operations automation in business by the exploitation of software system robots to cut back human efforts. Artificial intelligence robots can be used to measure the area of a property easily and accurately. This method may be primarily considered in a sequence of steps that result in meaty activity, Further in supervised learning, the algorithms consist of an end product which can be a variable, or this variable can be from a group of free-lance variables. Further by the exploitation of some operations, the input square measure portrays/portrayed the specified outputs, and some common couples of supervised learning are:

1. Regression
2. Random forest [5]
3. Call tree [6]
4. KNN
5. Provision egression

But in neglected learning, the algorithmic rule doesn't have any target variable which is used in the bunch for numerous teams. Here, a priori algorithmic rule K means principal element analysis square measure as some samples of neglected learning algorithms. Any method which is dispensed by any artificial machinery like the robots with minimal human intervention is known as automation and is closely associated with human ecology which focuses on rating prediction on the exploitation of the computer. This square measure is a spread of applications of machine learning like expressing emails whether it is tough to build a basic algorithm rule to perform the task effectively. Machine learning algorithms square measure altogether works/worked on data-set. It also measures the upper version of the regular algorithms to make the process "smarter" by allowing them to mechanically learn from the data-set given by the

users. In general, the property can have the valuation of land [7] for duty based to furnish. This will be a quantitative live off the profit claimed out by several different players in the industrial area for instant land agents. House price prediction is not that simple always, various because it includes a variety of middle men that influence the price of the final product. The property may work on the manners which exchange in the open market underneath several conditions and circumstances. AI will help customers to be more vigilant and thus help them save their hard earned money. This modification in the costs of properties shows an effect on each individual and also on the government to avoid such circumstances, and there's the desire for worth prediction.

II. RELATED WORK

Looking at the current scenario where urbanization is at its peak and people have largely started to move from villages to big cities, to facilitate a better life style. Thus expanding the real-estate market exponentially. It is very important to automate things, so that frauds in the name of property dealing could be controlled [7].

For every project, the literature review can provide a clear plan and its function with the bottom line, most of the authors have written over that artificial intelligence and neural networks which will influence and help making new and better algorithms for a better software that helps in prediction. By conducting this study it helps to grasp concerning the pros and cons.

The below passages describe the methodology utilized in the estate house value predictions and it has also used graphs and flowchart to explain certain concepts (Figure No :- 1 Graphs and flowchart to explain certain concepts in the estate house value predictions)

1.2 Data-set (Table No : 2)

House price prediction in urban areas:

In general, we are able to say that with the help of machine learning, we will be able to achieve new heights especially in terms of using historical data to make accurate predictions and be prepared for what is about to come next in our way dynamically, without being explicitly programmed on every case. Machine learning [8] focuses on the building method of computing primarily based programs that may access data-sets and use them to learn by the users.

2.1 Machine Learning and Its Challenges:

The self-learning of a computing system starts with information or data-set, like examples, direct expertise, or commands to see or search for patterns in data-set and build a decent function which

calls itself in the future supported examples. The essential goal is to allow machines like computer systems to learn from their own without any help from the human and modify certain actions in step with it.

2.2 Some Challenges that are faced in Machine Learning:

2.21 Inaccessible data and stored data Security:

Gathering data-set is not the whole concern. The collected data needs to be filtered to fit into the algorithm. Moreover with so much data in our database we always need to make sure that its 100% safe from any malicious attacks. Once a team or a corporation has worked on the data-set, safety may be a vital feature that has to be taken very seriously. Differentiating between the user's sensitive and insensitive data-set [9] is vital in implementing machine learning accurately and adeptly.

2.22 Infrastructure necessities for Testing & Experimentation:

The delaminate manner is typically accustomed to taking a look at machine learning algorithms. During this technique, the users are liable to draw an irregular sample from the data-set that may be a depiction of the actual population. The most common manner is to divide the data-set in an exceedingly delaminate fashion. Delaminate simply means the users need to split the data-set in order that every category is properly delineated within the ensuing subsets of the coaching.

2.23 Rigid Business Models

When the users implement machine learning it doesn't really mean the users will succeed. The method used to examine things is to be compelled if one plan isn't operating. For this, we need to make sure that we are operating and performing actions required quickly and also logically keeping in mind all the possible outcomes and the risk factors too.

2.24 Lack of Knowledge

With artificial intelligence and machine learning being new technologies within the information technology business, the talent pool needed to totally perceive and implement knotty machine learning algorithms is prohibited. We need a proper trained team of experts to take care of every aspect of the Machine Learning model being trained.

Organizations square measures reasonably realizing the approach to machine learning will open up for them. As a result, the demand for data-stable data-set scientists has aggrandized so have the salaries during this space. The job sites have listed the data-set scientist to be the best paying jobs out there. With heaps of organizations obtaining a smart beginning with massive data-set, artificial intelligence, and ma-

chine learning, the demand is going to increase heavily in the coming years.

2.25 Long Implementation

It includes heaps of complex computation and careful execution. And yet, thanks to multiple layers and the traditional unpredictability concerning the behavior of the algorithms, it's not sure that the time given by your team for machine learning project completion is going to be accurate. Therefore, it's vital to own restraint associated with an adaptation manner of doing wherever engaged in machine learning comes. To realize wished results on adoption, you are to offer enough time to your team and your project.

2.26 Affordability

Budgeting is a totally different milestone within the journey that works out the users to suit the affordability of the organization. If you're not assured of the talent needed in implementing the full-fledged machine learning formula, you'll be able to forever choose a consultation with corporations that have the experience and data-set when working with machine learning comes.

Nowadays things are shifting from analog or manual to digital or automated some examples of these are e-banking, e-learning, etc.

House Prediction Model Overview and Preliminaries:

In our well-planned system, the basic or primary step is to scrap data-set. it's a war with the assistance of that structured data-set is to fetch out from the net or any application and saved to info or programs or a CSV file or any other file.

The main aim of this project is to predict the house costs which helps in decreasing the issues faced by the client or vendor. The current manner or the analog way is that the client approaches a true land agent to manage his/her investments and discuss appropriate estates for his investments. However, this manner is sort of old and tough because the working-agent may predict the wrong worth of the property and so resulting in the loss of the customer's investment. The manual technique that is presently employed in the market is not that much helpful as per the current state and has vast risk. Therefore on getting the higher of this fault, there's a necessity for associate degree updated and strong system. Data processing algorithms[10] is ought to facilitate investors to take a position in associate degree sensible, these estates rely on their mentioned necessities. Additionally, the new system which the users are implementing can have a low price of working and can take less time as compared to the agents. This also has easy operations and a much elaborated way of choosing the estate in which the customer wants to invest. The planned system

works on regression toward the mean Algorithm largely.

Diagram:

Figure No :- 2 (Login and logout process)

Figure No:- 3 (Basic architecture design for house price prediction mode)

1. Pseudo code for the algorithm used in house price prediction

Input: x- the coordinate point to x

y- the coordinate point to y

Step-1. Read n //total number of points

Step-2. Read x, y //x and y co-ordinates of points

Step-3. First do Initialize diffx[n], diffy[n]

Step-4. Then Initialize diffxy, diffx2 to 0

Step-5. for i = 1 to n do

calculate the mean of x : xm mean of y : ym

diffx[i] = x[i] - xm //find the difference

values between

each x and mean of x

diffy[i] = y[i] - ym //find the difference

values between

each y and mean of y

diffx2 = $\Sigma(\text{diffx}[i])^2$ //calculate the sum-

mation of all the

difference values of x

diffxy = $\Sigma((\text{diffx}[i]) * (\text{diffy}[i]))$ //compute

the product

Of diff values of x and y

end for

Step -6. Find Slope using diffxy and diffx2

Step7. m = diffxy / diffx2 //the slope value is obtained by this

Step-8. c = ym - (m * xm) //the intercept value is obtained with this

Step-9 Equation complete: y = (m * x) + c

Step-10. Stop.

• Linear Regression:

Basically linear Regression technique [11]allows us to join and to learn about the link between two consecutive quantitative variables.

• The First variable that can be denoted as x is found to be the predictor, instructional, or irregular value.

• Second variable that can be denoted as y, is found to be responsive, resulting in irregular value.

• Multiple regression Analysis:

Basically, the case of Multiple regression is used to find if there's a scrupulously notable union in the center of the group of values or not. Multiple regression usually found patterns in different recurring exploration are going to be extremely close. As a simple straight re-occurrence. The highest value differentiate is in the mid of basic linear re-occurrence and in addition to re-occurrence in between the variants for sug-

gestions (“x” variables) used within those relapses. Easy Reoccurring scrutiny employments associate absolute x variable to each subordinate “y” variable. Case in point: (x1, Y1).

Many reoccurring usages varied “x” variables for every free var: (x1)1, (x2)1,(x3)1, Y1). In the first var straight regression, you would find in all probability of data-set indicating case subordinate variable (i. E. “gain”). Anyways you will make fascinated by how-ever different types of forms maintaining offers impact the re-occurrence.

• **the value perform:**

In value perform generally the users enlarge the scale of a chosen search, basically the space you anticipate that those deals would probably an opportunity to be elevated. Despite escalating those sizes, those negotiations in this don’t seem to elevate that a sublime deal. Almost those outlay connected before, elevating those span of the design, on the condition for you negative results. So, the users have a proclivity to prerequisite on cut back on these prices. The users have a proclivity to present accomplice in health cost perform, that is vitally familiarized characterize and take into account those slip of the model.

$$\text{Mean Square Error} = \frac{1}{n} \sum_{i=1}^n (Y_i - Y_i')^2$$

Mean square error:

• **Lasso Regression**

In the case of Lasso regression which may be accomplice uncustomary between those deteriorated models that might approachable can fetch the data-set from CSV file. After the regression model as well as exemplify for a sample, the method is in computation recorded to recommend. LASSO refers to the least fixed change and chosen Operator.

Lasso regression is one of the sparse among the standardization patterns that form mean models among the local for a large no. for features, the site of extravagant instigate to which the consecutive two objects.

• Huge enough to recreate those ramp-up of the model on over-fit and at least around ten var can be found under the category of overfitting.

• Large enough will lead to various tests. This state of the affair would probably come out among the occurrence from affirming link in Nursing oversized selection or billions regarding features. Whether re-occurrence performs L1 standardization as it involves those batch measures equal to the foremost admiration of the range of the values. In simple words, here the cutback objectives are regarding demonstration surpassed.

Minimization goal = LS Object + λ (sum concerning outright esteem of coefficients).

The place LS Object remains for lowest squares objective that is in the position to be nothing yet the straight re-occurrence of a target whereas not standardization is more of addition to those changing figures that controls the life for standardization. The propensity would build with those enhancing the quality of λ and thus the eminence will reduce regarding engraving the life for shrinkage (λ) improvements.

Here the turning part controls those featured out for sanction. Once the users have a propensity to persuade the same concomitant equally as basic right re-occurrence. At λ = ∞: constantly on coefficients unit of evaluation zero. The objective once zero < λ < ∞: the users incline to persuade coefficients between zero what's tons of that for basic straight re-occurrence thus at λ is midst the two extremes, we'd adjusting those below two plans.

- Fitting link in Nursing right model for y once X.
- getting those coefficients.

Gradient Boosting:

Gradient boosting may machinery dependent to maintain the plan of action to reoccur together marshaling complications, that build a prognostication model between the order of link in health cluster from

affirming helpless prognostication models. The truth of a prognosticator model is to most likely be grasping features building alternately. Toward implanting improving computation rightmost. There measuring unit is a giant vary improving computation.

In the case of Gradient Boosting may be a machine learning algorithm for regression and classification complications that generates a prognostication model among the design of assistant in the nursing troupe of weak prognostication models, generally decision trees.

- **XG Boost:** Basically in the case of XG Boost, an assistant in Nursing improves scattered gradient by improving the library, mapped out to be very inexpensive, flexible, and adjustable. It applies machine learning algorithms under the Gradient improving framework. XG Boost gives a parallel tree improvement that resolves huge data from the data-set solving complications in the link of nursing passing in a regular manner. The same instructions run on maximum dispersed atmosphere (Hadoop, SGE, and MPI) and should resolve complications so as much aspect a huge number of exemplars.

- **AdaBoost:**

It is a link in Nursing with whole learning (also known as “meta-learning”) that was at the starting formed to enhance the order of binary analysis. It uses the link in nursing monotonous approach to finding out the mischiefs of feeble classifiers, and change them into long-lasting ones.

- **Mild Boost:**

In artificial intelligence, every single improving algorithm wants its own primary computations. Also, a sudden selection is seen at the same time when applying them. Improving computation are about to be unusual among those. The larger 0.5 i.e. in an improving position taking in notation oriented among the ideal one twenty a drawn-out time. It would lead to arrange complications it will be developed to re-occurrence. The inventiveness of gradient improving may have been a link in the nursing facility. It links those o/p regarding huge 0.5 “

Weak analyzer to form a proficient “panel.” a helpless classifier (e. G. choice tree) is about to be whose slip rate primary valuable to irregular dead reckoning

III. IMPLEMENTATION:

3.1 Experimental Result and Discussion:

The experimental result shows that the planned system data-set holds some anomalies before as the users will see here that 2bhk flat with same data-set showing additional worth than the 3bhk data-set with same specifications. Figure No :- 4 (experimental result 1)

Later on when the users applied some outliner to removal function the users get some perfect re-

sult to predict price on more suitable data and the final result data can is shown in scattered graph below . Figure No :- 5 (experimental result 2)

IV. CONCLUSION AND FUTURE WORK:

Here the users have managed to prepare a robust model that can predict the perfect value for a real estate company to compete. It provides costumers or users the best way to take a glimpse at future advances, the value of prognostications. A few worsen ways have also been found. Furthermore, when arriving during a prediction way in light of XG support. Straight former imply works been utilized within our model, something like that the future value predictions will have a capability towards all the sensible values. The users prepared an approach with the use of similarly as noticeably information as time enables for our prediction system, by adopting those ideas from claiming ramp boosting.

Include All the functionalities under the model, the users can give an alternative for the client like select a district alternately locals should generate those high-temperature maps, as opposed to entering in the list.

REFERENCES

- [1]. Sifei Lu, Zengxiang Li, Zheng Qin, Xulei Yang, Rick Siow Mong Goh, “A Hybrid Regression Technique for House Price Prediction”, December 2017.
- [2]. Ayush Varma, Abhijit Sharma, Sagar Doshi, Rohini Nair, “House Price Prediction Using Machine Learning And Neural Networks”, INSPEC number 18116205, April 2018.
- [3]. Adyan Nur Alfiyatin, Hilman Taufiq, Ruth Ema Febrita, Wayan Firdaus Mahmudy, “Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 10, 2017.
- [4]. T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, “How many trees in a random forest?” In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7376 LNAI, 2012, pp. 154–168, ISBN: 9783642315367. DOI: 10.1007/978-3-642-31537-4\13.
- [5]. S. Ray, “CatBoost: A machine learning library to handle categorical (CAT) data

- automatically,” CatBoost: Analytics Vidhya, 14-Aug-2017.
- [6]. Neelam Shinde, Kiran Gawande, “Valuation of House Price Using Predictive Techniques”, International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835(IJAECs), Volume-5, Issue-6, June-2018.
- [7]. Eli Beracha, Ben T Gilbert, Tyler Kjorstad, Kiplan womack, "On the Relation between Local Amenities and House Price Dynamics", Journal of Real estate Economics, Aug. 2016.
- [8]. R. J. Shiller, “Understanding recent trends in house prices and home ownership,” National Bureau of Economic Research, Working Paper 13553, Oct. 2007. DOI: 10.3386/w13553.
- [9]. S. C. Bourassa, E. Cantoni, and M. Hoesli, “Predicting house prices with spatial dependence: a comparison of alternative methods,” Journal of Real Estate Research, vol. 32, no. 2, pp.139–160, 2010. [10] Li, Li, and Kai-Hsuan Chu. "Prediction of real estate price variation based on economic parameters." Applied System Innovation (ICASI), 2017 International Conference on.IEEE, 2017.
- [10]. Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." Journal of machine learning research 12.Oct (2011): 2825-2830.
- [11]. Byeonghwa Park , Jae Kwon Bae (2015). Using machine learning algorithms for housing price prediction , Volume 42, Pages 2928-2934. [13] Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Visning, 2015. Introduction to Linear Regression Analysis.

Real-Time Intelligent Heterogeneous Networks - Applications, Challenges, and Scenarios in IoT HetNets:

Figures And Tables

Figures:-

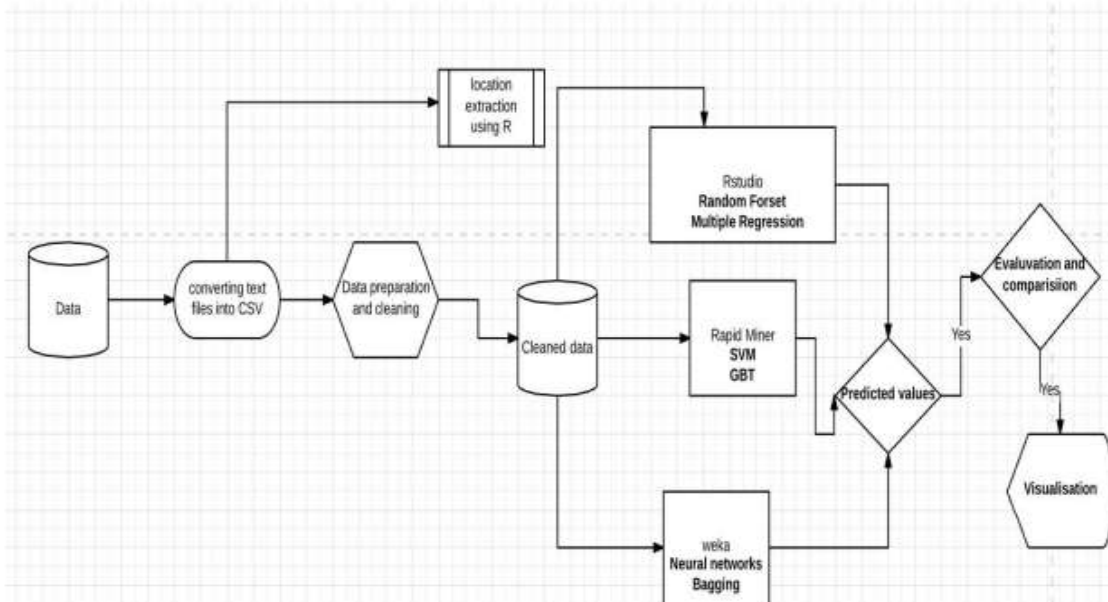


Figure No 1 (Graphs and flowchart to explain certain concepts in the estate house value predictions)

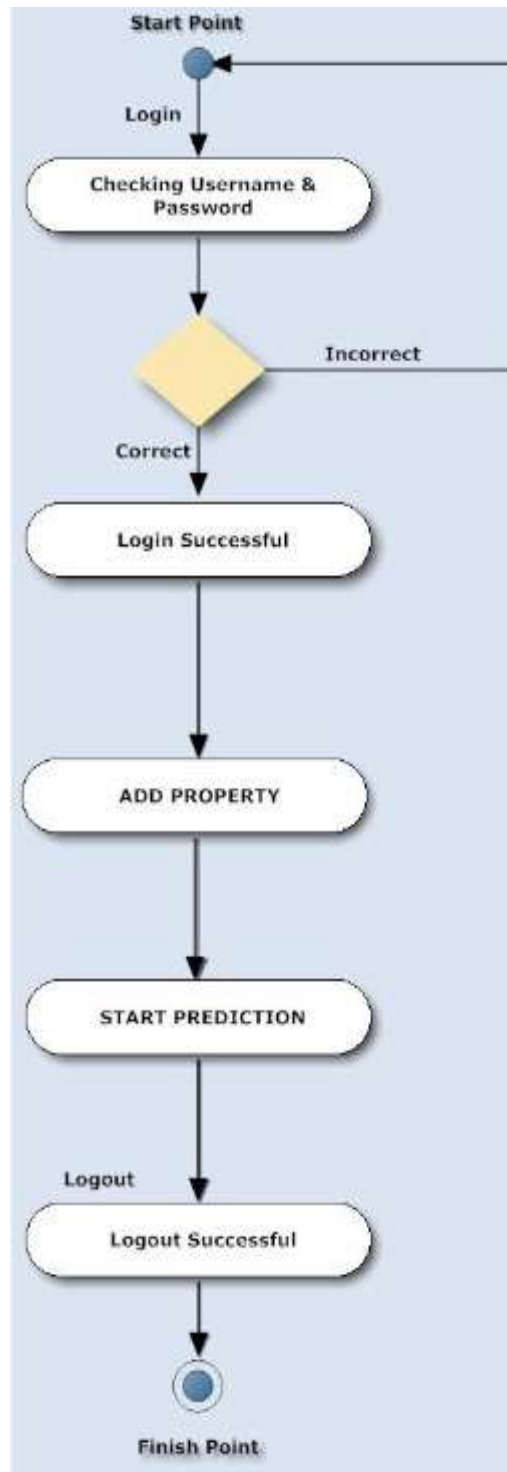


Figure No :- 2 (Login and logout process)

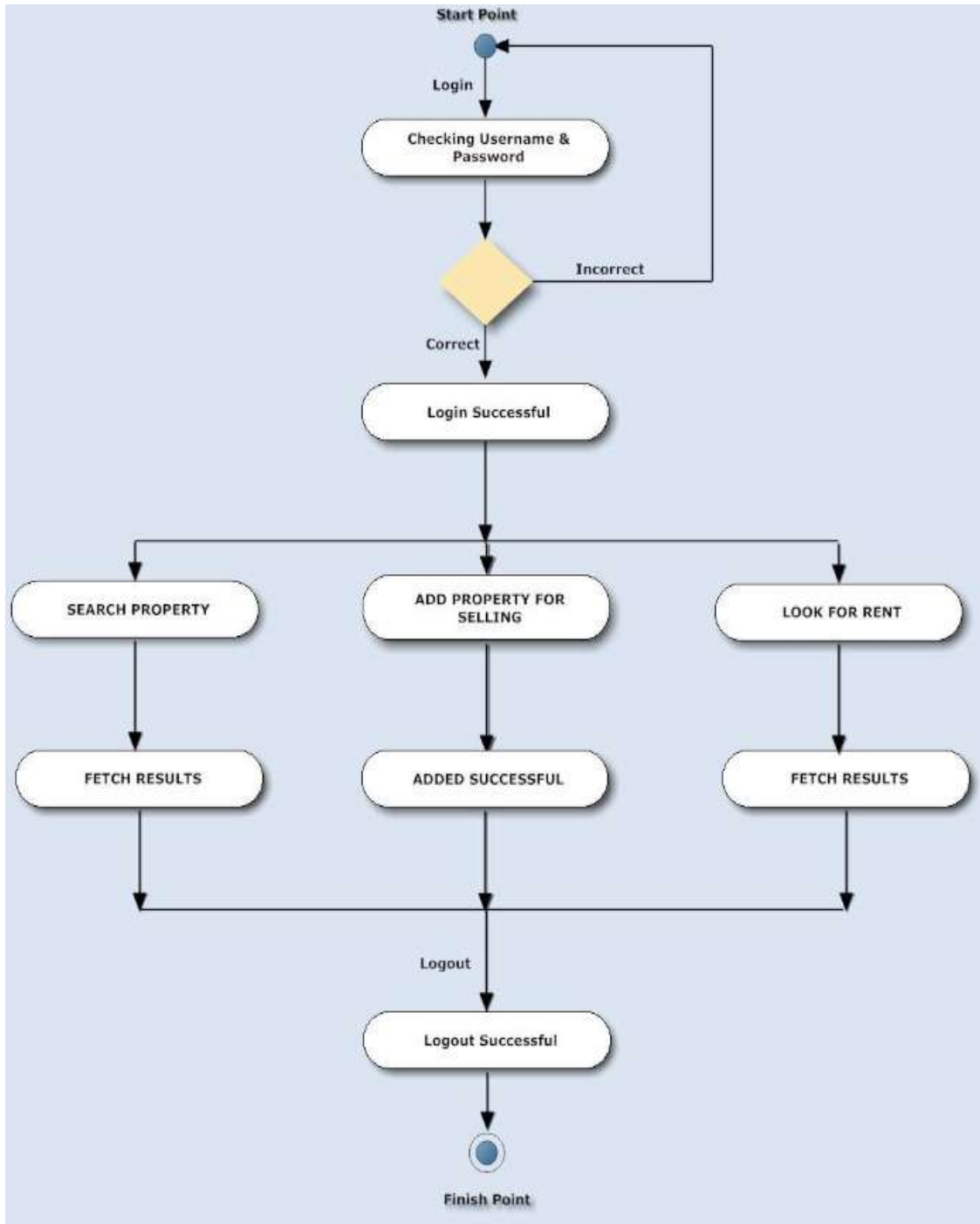


Figure No:-3 (Basic architecture design for house price prediction mode)

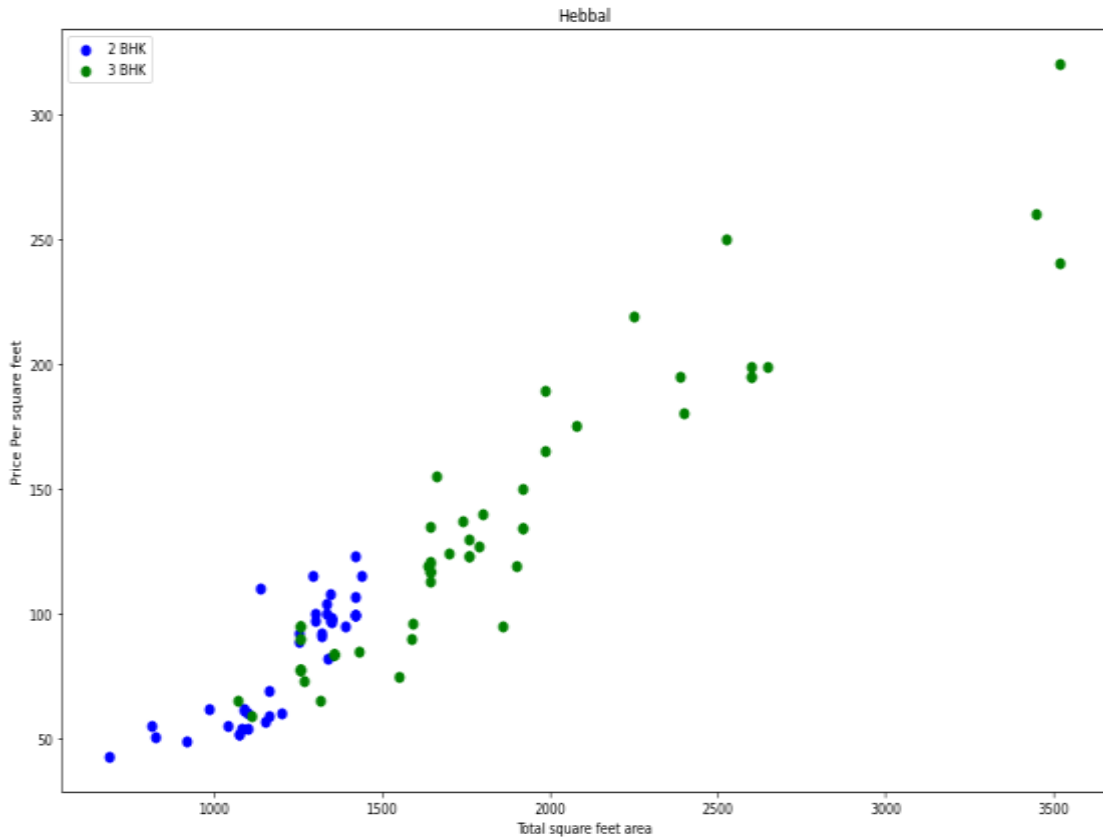


Figure NO :- 4 (experimental result 1)

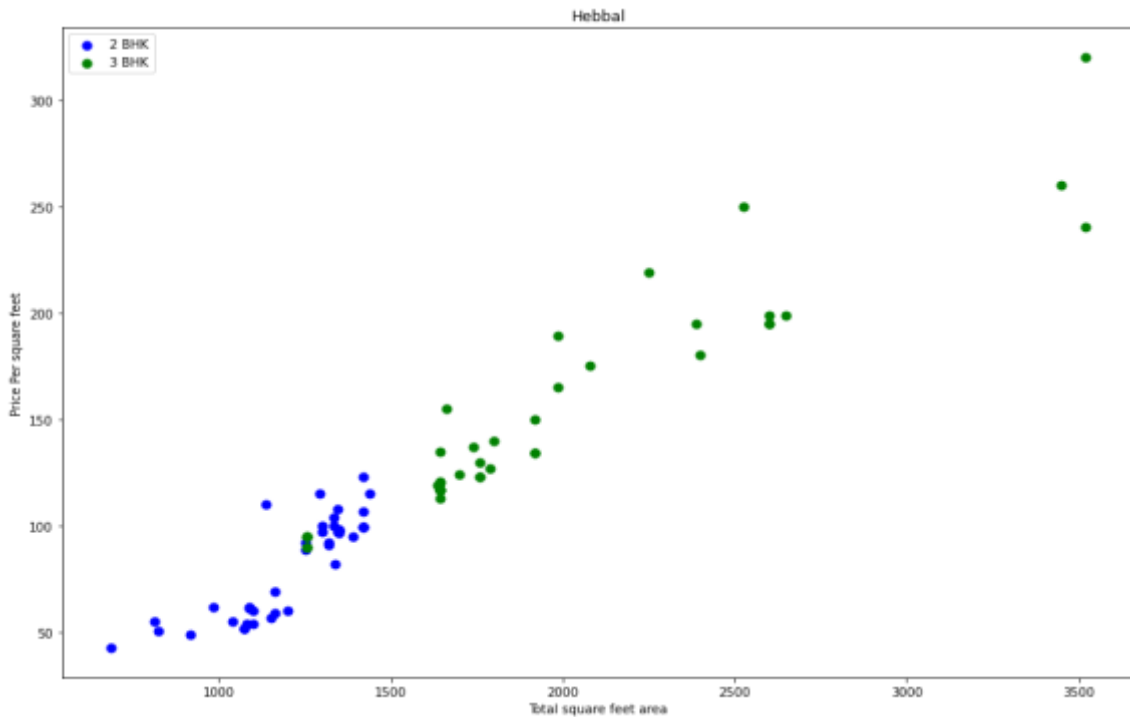


Figure No :- 5 (experimental result 2)

Tables

S.no	Variables	Lowercase
1	location	String
2	size	Integer
3	Total_sqft	Float value
4	bath	Integer
5	price	Integer
6	bhk	Integer
7	price_per_sqft	Floating value

Table No :-1(Variables,lowercase)