Predicting property prices- A Universal Model

By

Mayank Kumar Nagda

(RA1511003010313)

Under the guidance of

Dr. E. Poovammal

A Minor Project Report

Submitted to the Department of Computer Science and Engineering

In partial fulfillment of the requirements
for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

of

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM UNIVERSITY
(Under section 3 of UGC Act 1956)

S.R.M. Nagar, Kattankulathur, Kancheepuram District

October, 2017

Predicting property prices- A Universal Model

By

Mayank Kumar Nagda

(RA1511003010313)

Under the guidance of

Dr. E. Poovammal

A Minor Project Report

Submitted to the Department of Computer Science and Engineering

In partial fulfillment of the requirements
for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

of

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM
UNIVERSITY
(Under section 3 of UGC Act 1956)

S.R.M. Nagar, Kattankulathur, Kancheepuram District

October, 2017

SRM UNIVERSITY

BONAFIDE CERTIFICATE

Reg No: RA1511003010313

Certified that this project report titled ***Predicting property prices- A Universal Model*** is the bonafide work of ***Mr.Mayank Kumar Nagda.*** who carried out the research under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Signature of the Guide   Signature of the H.o.D

Name of the Guide

Date:

Abstract

The goal of this project is to develop a universal model for predicting property prices of any region given detailed information. The model will be useful for many purposes, from estimating the worth of a property that's not on market, to figuring out which component factors the most in buying a property. Developed model helps in deciding whether to invest for remodeling property or not.

Property prices are important reflection of economy. Price ranges are of great interest for both buyers and sellers. In this project, property prices will be predicted given explanatory factors that cover many aspects of residential properties. As continuous values, property prices will be predicted using Linear Regression and Clustering will be used to make a generalized universal model.

The aim of this project is to develop a universal prediction model for property rates. Taking into consideration that this model is applicable for any region universally, the accuracy may be compromised for some areas initially. Applying suitable Machine Learning algorithms guarantees improved accuracy with every prediction.

Acknowledgement

In completing this Minor Project, I have been fortunate to have help, support, and encouragement from many people. I would like to acknowledge them for their cooperation.

First, I would like to thank Dr. E. Poovammal, my project guide, for guiding me through each and every step of the process with knowledge and support.

As every great achiever is inspired by a great mentor, I would also like to take this opportunity to thank my Faculty Advisor Ms. P. Mahalakshmi for her constant support, inspiration, and guidance from the very start.

I would also like to thank Dr. Annapoorani Panaiyappan, my Academic Advisor who helped me throughout the development of the project.

It is my radiant sentiment to place on record my best regards, deepest sense of gratitude to faculties of SRM University to include a Minor Project course in our curriculum, which pushed me to take this opportunity which was a great experience for learning and professional development.

I perceive this opportunity as a big milestone in my career development. I will strive to use gained skills and knowledge in the best possible way, and I will continue to work on their improvement in order to attain desired career objectives.

Sincerely,
Mayank Kumar Nagda
RA1511003010313

TABLE OF CONTANTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, ABBREVIATIONS OR NOMENCLATURE

- *A1, A2, A3…. An – Attribute 1, Attribute 2, Attribute 3…. Attribute n*
- *P1, P2, P3…. Pn – Priority Points 1, 2, 3 .. .. Priority Point n*
- *W1, W2- Weight Point of Property 1, Property 2*
- *S1, S2- Selling Price of Property 1, Property 2*

1. INTRODUCTION

Property rates of any region are always a topic of discussion within the society. There are many factors which affects property rates for a region. Also, these factors might not be same for all the locations. In real world there are constant negotiations while purchasing a property. Hence, there can only be a range of actual property rate and not the exact rate itself.

1.1 EXISTING MODELS

The quest for predicting property rates has been studied for a longtime in many fields, including statistics, patterns recognition and exploratory data analysis. Analyzing the explicitly stored data provides further knowledge about a business by going beyond the data to derive knowledge about the business.

Most of the prediction models [1][2][3] developed so far have a common way of collecting all the property rates in a given region. Then applying a suitable mathematical model to generalize the collected data for that region, helps in prediction (rates).

This type of prediction model entirely depends on the mathematics used and also, they hold true for a particular region where rates have fewer variations. These models can’t be used in places with drastic variations in property rates.

A universally generalized model is one which can be used as such in any condition (region) and is flexible enough to adopt itself according to the conditoins. While doing so, model also should hold the competitive mathematical success.

1.2 PROPOSED MODEL

To generalize any model universally, some factors need to be considered which affect the property rates of that region. As discussed in section 1.1, these factors may not be same for all the regions. A generalized model is one which is flexible enough to change itself with any particular region according to its factors.

Property rates for any region are never constant. They keep on changing with the market transition. The generalized model should also be intelligent enough to change itself with the market growth. What is the use of a model which gives outcomes of the past? A real prediction model is one which can predict the future property rates and trends.

Once generalization problem is solved, different mathematical techniques such as Linear Regression, Clustering can be applied to predict property rates.

2. LITERATURE SURVEY

The biggest problem faced in developing a universally acceptable property prediction model is the lack of property datasets. It's not possible to have data set of property prices of the whole world. Therefore, most of the prediction models are limited to specific regions only.

Techniques such as Linear Regression with Support Vector Regression (SVR), k-Nearest Neighbors (kNN) and Regression Tree/Random Forest Regression were used by **Nissan Pow, Emil Janulewicz, Liu (Dave) Liu [1]** to predict the prices in Montreal (Canada). Very large preprocessed datasets were used in this project. It successfully predicted both the asking and sold prices of real estate properties based on the features such as geographical location, living area etc.

In this model [1] the input data is main housing prices of the whole municipality of Montreal. This data is used in mathematical model to carry out predictions. This model failed to encounter the rate change factor, which often happens quarterly as the prices go up or down.

Because the dataset was taken directly, the real world perception of people living in Montreal, who actually buy the property was not considered. And also, the model [1] is only focused and limited for Montreal, Canada. The same model wouldn't give the same accuracy if used for any other city.

A prediction model for real estate customers developed by **Nihar Bhagat, Ankit Mohokar, Shreyansh Mane [2]** has a basic objective of reducing human factors which always play a part for sharing the views on which property to purchase and also to provide information about the future market trend for that property.

In this project [2] Data Mining and Machine Learning are combined for predicting property rates. The model uses Linear Regression algorithm. This paper efficiently analyses previous market trends and price ranges, to predict future prices. This topic brings together the latest research on prediction markets to further their utilization by economic forecasters.

It [2] doesn’t predict future prices of the houses mentioned by the customer. Due to this, the risk in investment in an apartment or an area increases considerably. To minimize this error, customers tend to hire an agent which again increases the cost of the process. This leads to the modification and development of the existing system.

Aaron Ng of Imperial College of London [3], also tried a hand on this matter but was in need of very large dataset and also the model was limited to the properties in London. He developed an Android application that provides users with information on the geographical variations of future London housing prices through heat maps/graphs. He harnessed the power of Distributed Gaussian Processes as a machine learning technique to build the prediction system and maximize the dataset utilization.

For the model [3] he used around 2.4 million datasets to carry out this operation and still that was limited to only London.

2.1 LIMITATION IN EXISTING MODELS

The real problem lies in the market itself. Property rates keep on changing and there are too many factors affecting it. In some area, it’ll be a new supermarket which resulted in high property rates. Somewhere new facilities are coming which again result in increasing rates. It becomes very tough to predict property rates with a model which is 1-2 years old because here, the treads are changed daily.

Now with these many factors and problems in the way, most of the researchers limit their model to a particular location, ex. London, Boston, Montreal etc... But what about other places? This can only be done with one generalized model as discussed in 1.2.

3. GENERAL TRENDS IN PROPERTY PRICES AND PROPOSED MODEL

Trends in property rates change very fast. A property lying dead somewhere in a corner of a city can suddenly become a hotspot if there is any market change nearby, such as the opening of new supermarket or university in that area. These factors drastically affect the property prices.

3.1 Factors Affecting Property Prices

Property rates in any given area depends on the factors-

1. Location
2. Supply and Demand
3. Market/Economic growth
4. Condition
5. Neighborhood

These factors vary for different areas. Property rates of a good location in a normal city differs from that of a metropolitan city.

Globally there can’t be anything such as a good location or a bad location. It always depends on the place relating about.

Same thing goes for all the factors. A good neighborhood in Texas has a very different meaning than that of in Miami.

3.2 A Generalized Model for Predicting Property Rates

All the factors which affect the property prices are relative from location to location and so is their fundamental meaning. Therefore, it’s not possible to have a generalized model which has some fixed values to predict the property rates.

The solution to this problem is to divide different location with varying prices into different clusters. In these clusters, the factors which affect the property prices will have same meaning. The motto is to prepare a universal model which can predict the property rates in these individual clusters.

Final Model comes in action when after combining all these clusters into one and get one single contour solution.

Figure 1 is an example clustering model, formed by random datasets. The bar shows property prices and the region within shows the resulting cluster formed by it.

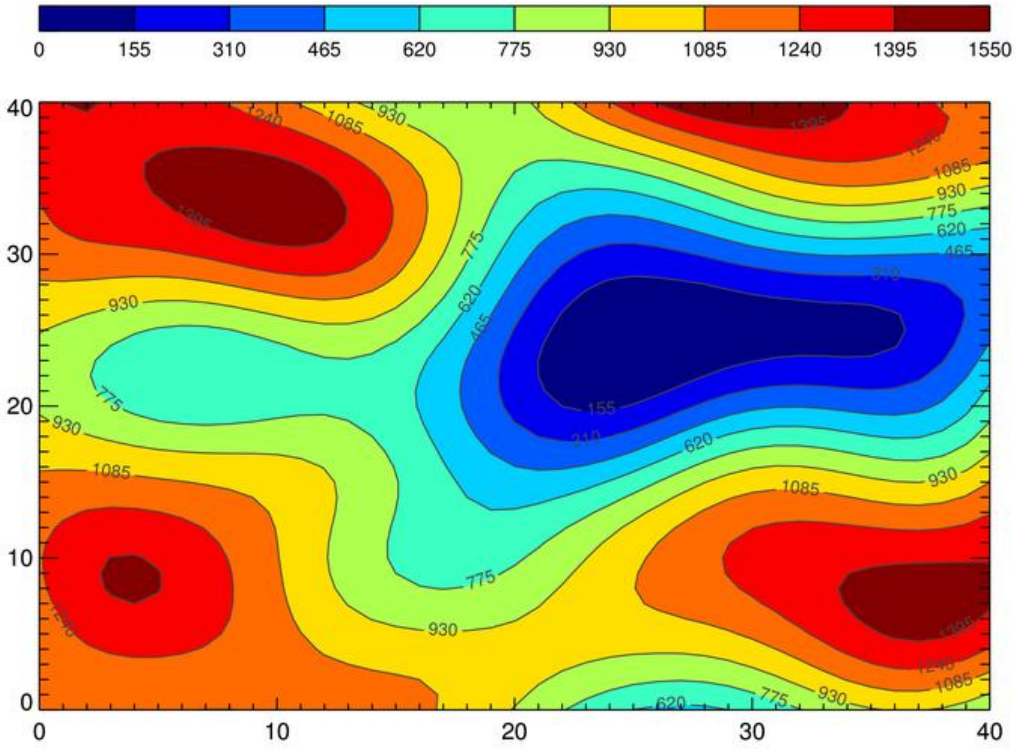


Fig. 1 Clustering Contour

Following points can be inferred from the discussion.

1. Area of any size can be divided into clusters.
2. Clusters are formed after calculating prices in small regions and then combining them.
3. From the figure, it can be depicted that warm colors represent higher property rates and cold colors represent lower property rates.
4. The weight of each cluster is directly proportional to the price it carries.

3.3 Cluster Model

Knowing the target model and to get the desired predictions, custom datasets of property prices in a particular region are calculated. Adding these regions, desired cluster model is achieved with varying property prices in different regions.

3.4 Regional Prices and factors affecting them

Regional prices also follow the same factors as discussed in 3.1 but the difference being that it can now be addressed. It’s possible to find good neighborhood (and other factors) in a particular selected region because that region is now addressed separately.

As factors can now be addressed, every region has its own quality or depending factor. The target now is to survey this region and find desirability of the location from the regional people. The more desired the area is, the costly it’ll be.

4. PREDICTING REGIONAL PROPERTY RATES-

A NEW APPROACH

In this section, a new approach to predicting regional property rates is visualized, which is exclusively dependent on a particular region as well as can be further changed with time when the trends shift.

4.1 Surveying Regional Factors

As every region has different factors affecting its property price, it's best to survey it. A region is now surveyed and the desirability attributes according to people in that area is given by-

Survey 1:

A1 – Attribute 1

A2 – Attribute 2

A3 – Attribute 3

A4 – Attribute 4

.

.

.

An – Attribute n

These are the places *(A1….An)* in that region where people want their house closer to, but the preferences also differ from person to person, so to make it more accurate one more survey is conducted to ask people about their priority point given to that particular attribute/place.

Survey 2:

P1- Mean priority points for Attribute 1

P2- Mean priority points for Attribute 2

P3- Mean priority points for Attribute 3

P4- Mean priority points for Attribute 4

.

.

.

P5- Mean priority points for Attribute 5

On the basis of both the surveys, attributes as well as their priority points *(P1….Pn)* are collected which people prefer in that region.

4.2 Prediction Algorithm

A prediction algorithm is formed which uses the survey data collected in section 4.1-

Line 1: *Select a property in a region for which prediction is to be done.*

Line 2: *Get selling price (S1 & S2) of any two properties in that region which were sold in last few months.*

Line 3: *For each of the two selected properties, calculate their weight points (W1 & W2).*

To calculate weight points-

Step 1: Calculate distance of those two properties from the Attributes (A1... An)

Step 2: Multiply (10-distance) to Priority points of each Attributes.

*Step 3: Weight point is sum all such attribute * priority points.*

Ex. Individual point of each Attribute = (10-distance)

Weight Points = ∑ (individual point of each Attribute) x
(priority point of that Attribute)

Line 4: *Dataset now has weight points of two properties (W1 & W2) with their selling price (S1 & S2).*

Line 5: *Calculate weight points of preferred property (W3) and put that in linear regression with W1, W2, and S1 & S2 to calculate S3 (selling price of the selected property).*

This might seem a bit confusing and complex at one go but with the next chapter as implementation, it’s explained clearly with an example.

5. CONSTRUCTING A PREDICTION MODEL FOR PROPERTY RATES

Implementation of the proposed model includes a website, which takes the required survey data discussed in section 4.1 as input from the user and then predicts the outcome.

5.1 DESIRABILITY OF PROPERTY IN A REGION

An open survey was carried out to find which places *(Attributes)* regional people want near their house, or what effects property prices in their region.

Table 1 shows most common places which people preferred for their residential properties-

Table 1 survey of places effecting house prices

Places
Office/College/School
Bus Station
Railway Station
Super Market/Shopping Complex
Park/ Places to chill
Good Restaurants
Hospital
Places of Religious Importance

From the survey, it made clear what are all the places having their effects on the residential property prices. Another open survey is to be conducted, which asks for individual priority points of all the places.

Table 2 shows the Places and Average priority points it got based on the conducted survey.

Table 2Table having places with their priority points(out of 10)

Places	**Average priority points (out of 10)**
Office/College/School	6.2
Transportation Facility (Bus/Rail/Air)	5.8
Super Market/Shopping Complex	9.1
Park/ Places to chill	8.5
Good Restaurants	8.5
Hospital	8.9
Places of Religious Importance	5

5.2 Developing a User Interface

For the model, user input is required for data of two neighbor properties. An interactive UI system needs to be developed which asks for the input of the required data and sends it for the back end manipulation.

5.2.1 A simple GUI

For this project, an interactive website is developed by considering the popularity of internet these days.

Figure 2 shows the front page of the website. The sections on the website page are - About, How and Check It to check the property prices. Figure 4 shows front end code snippet.

Figure 3 exclusively shows Check It section which asks for input data from the users.

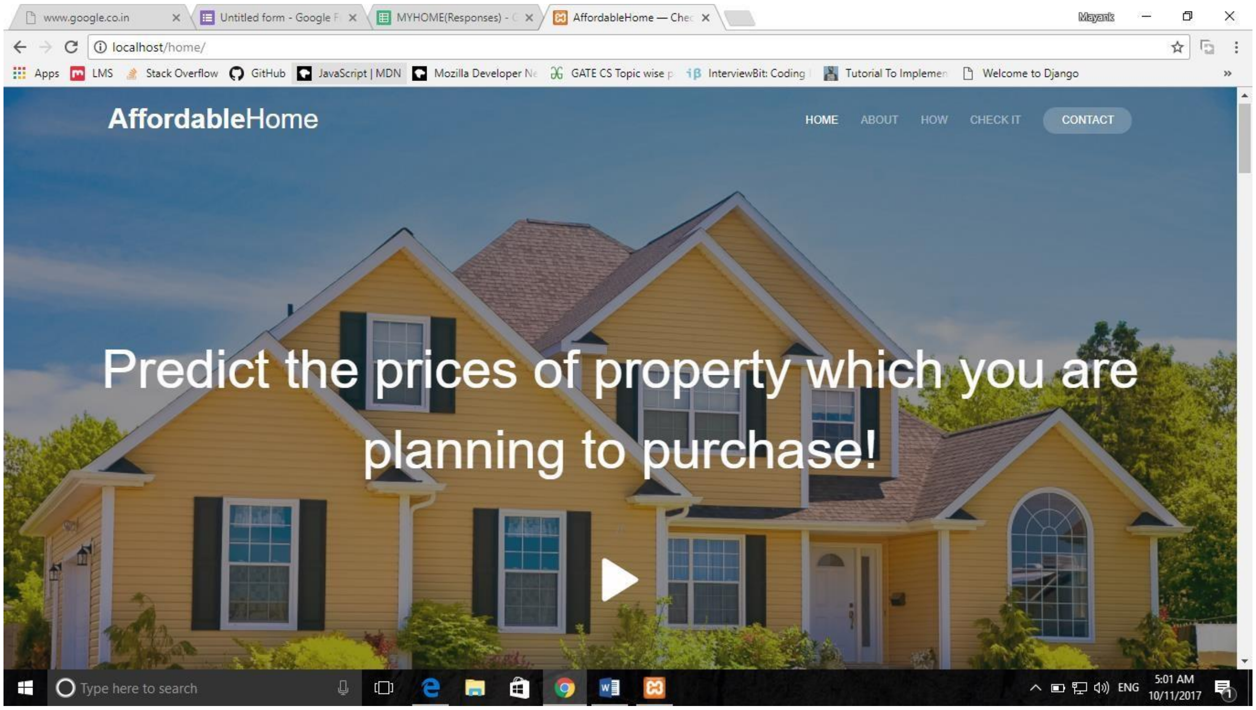

localhost/home/
AffordableHome
HOME
ABOUT
HOW
CHECK IT
CONTACT
Predict the prices of property which you are planning to purchase!

Fig. 2. Websit snap(UI)

To create this graphic UI, these are the technology

used-

- HTML
- CSS
- JavaScript

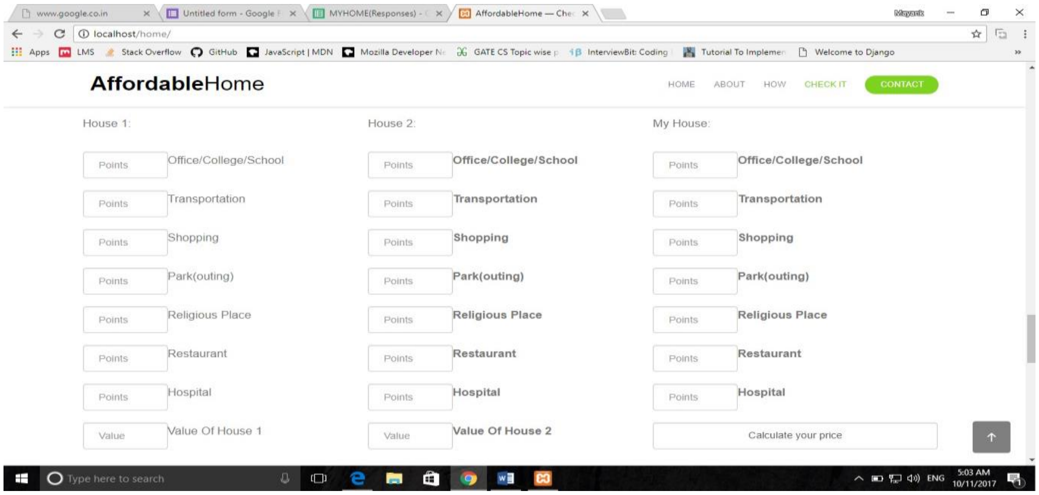
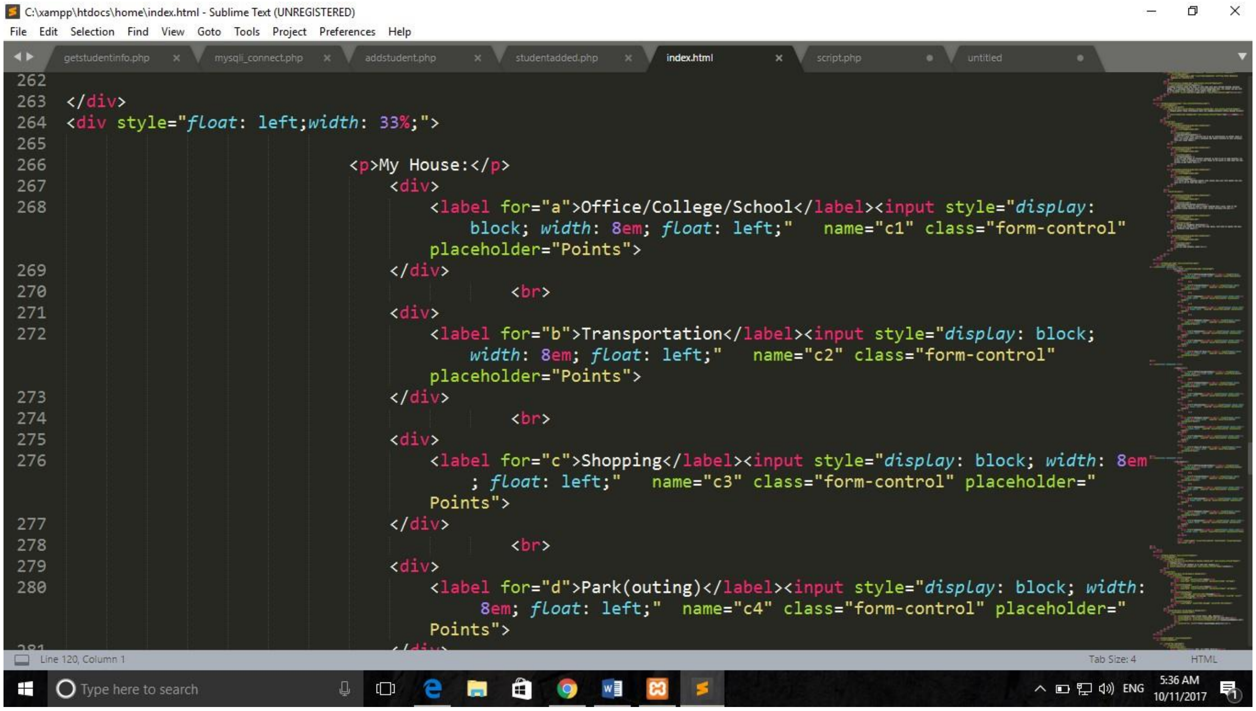


Fig. 3 User Data Input



```html

</div>
<div style="float: left;width: 33%;">

                            <p>My House:</p>
                                <div>
                                    <label for="a">Office/College/School</label><input style="display:
                                        block; width: 8em; float: left;"   name="c1" class="form-control"
                                    placeholder="Points">
                                </div>
                                            <br>
                                <div>
                                    <label for="b">Transportation</label><input style="display: block;
                                        width: 8em; float: left;"   name="c2" class="form-control"
                                    placeholder="Points">
                                </div>
                                            <br>
                                <div>
                                    <label for="c">Shopping</label><input style="display: block; width: 8em
                                        ; float: left;"   name="c3" class="form-control" placeholder="
                                    Points">
                                </div>
                                            <br>
                                <div>
                                    <label for="d">Park(outing)</label><input style="display: block; width:
                                         8em; float: left;"  name="c4" class="form-control" placeholder="
                                    Points">
```

Fig. 4 Front End code snippet

5.3 Data Management

For the backend development, php, MySQL are incorporated in the HTML using Apache, MySQL servers.

5.4 Execution of the Model

Based on the Algorithm discussed in Section 3.2 the execution steps for this model are:

Step 1:

Input is taken of every property for 7 different fields, each field deciding how far that place is from the house (in km).

$$\textit{Individual point of each field} = (10\text{-}\textit{distance})$$

Line 2:

Individual points of each property =

$$\sum \textit{(individual point of each field)} \times \textit{(priority point of that field from Table2)}$$

Step 3:

Individual points of each property with its price (from user) is now known.

Step 4:

This information in now stored in database systematically.

Step 5:

There is a possibility that database system already has previous datasets entered by some other user, so the whole dataset is retrieved to predict the price of the 3rd property.

Step 6:

The complete dataset is passed through linear regression algorithm to calculate slope and intercept.

Step 7:

After calculation of slope and intercept price is predicted by-

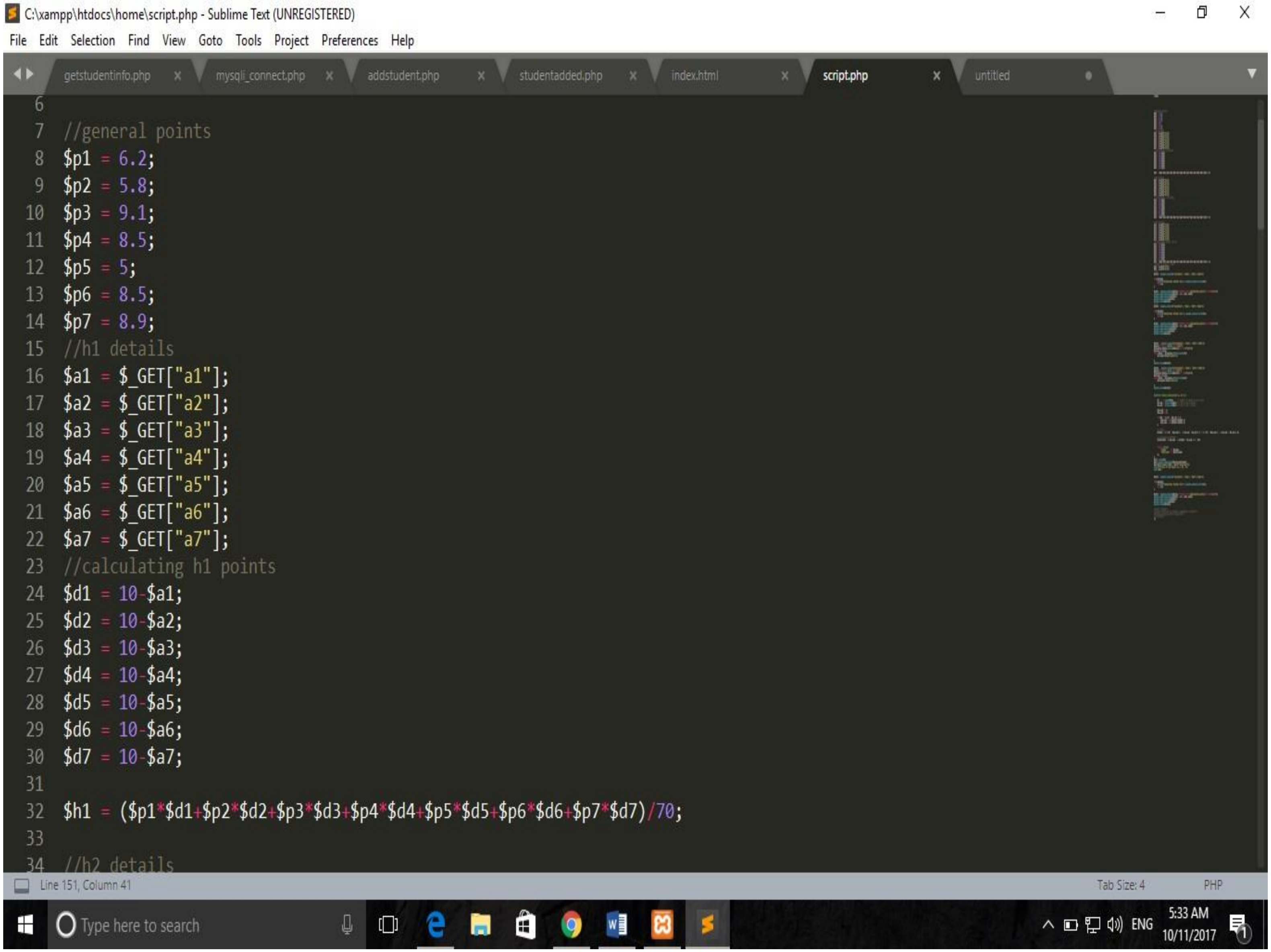
$$y=mx+c$$

y=y-axis element; x= x-axis element; m =slope; c= y-axis intercept

Step 8:

The data input, as well as the data predicted is again stored so that it can be used by some other user, and that's how it'll improve itself with every prediction.

Figure 5 and Figure 6 show some backend code snippets for reference.



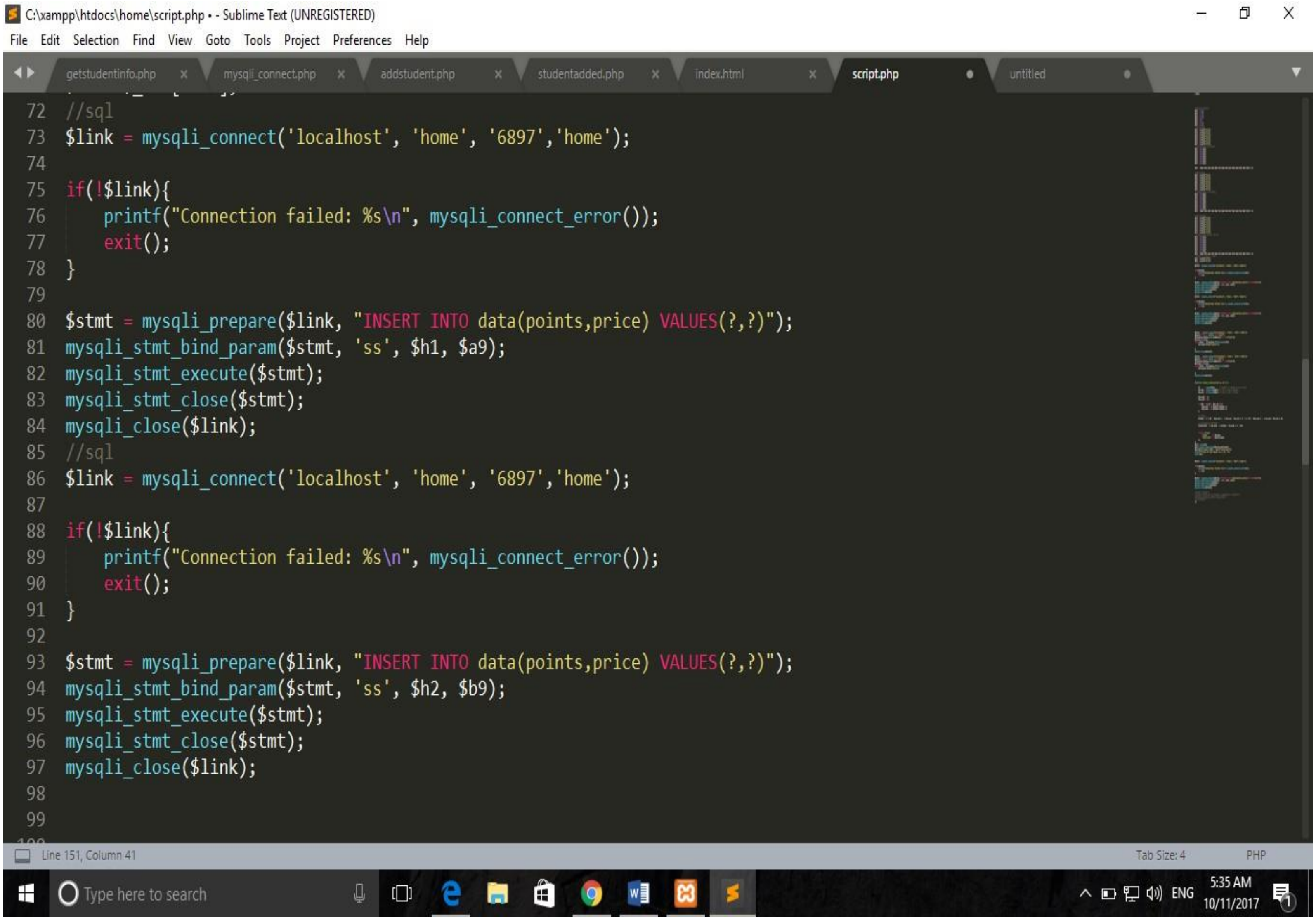
```php

//general points
$p1 = 6.2;
$p2 = 5.8;
$p3 = 9.1;
$p4 = 8.5;
$p5 = 5;
$p6 = 8.5;
$p7 = 8.9;
//h1 details
$a1 = $_GET["a1"];
$a2 = $_GET["a2"];
$a3 = $_GET["a3"];
$a4 = $_GET["a4"];
$a5 = $_GET["a5"];
$a6 = $_GET["a6"];
$a7 = $_GET["a7"];
//calculating h1 points
$d1 = 10-$a1;
$d2 = 10-$a2;
$d3 = 10-$a3;
$d4 = 10-$a4;
$d5 = 10-$a5;
$d6 = 10-$a6;
$d7 = 10-$a7;

$h1 = ($p1*$d1+$p2*$d2+$p3*$d3+$p4*$d4+$p5*$d5+$p6*$d6+$p7*$d7)/70;

//h2 details
```

Fig. 5 Code Snippet1- Backend



```
//sql
$link = mysqli_connect('localhost', 'home', '6897','home');

if(!$link){
    printf("Connection failed: %s\n", mysqli_connect_error());
    exit();
}

$stmt = mysqli_prepare($link, "INSERT INTO data(points,price) VALUES(?,?)");
mysqli_stmt_bind_param($stmt, 'ss', $h1, $a9);
mysqli_stmt_execute($stmt);
mysqli_stmt_close($stmt);
mysqli_close($link);
//sql
$link = mysqli_connect('localhost', 'home', '6897','home');

if(!$link){
    printf("Connection failed: %s\n", mysqli_connect_error());
    exit();
}

$stmt = mysqli_prepare($link, "INSERT INTO data(points,price) VALUES(?,?)");
mysqli_stmt_bind_param($stmt, 'ss', $h2, $b9);
mysqli_stmt_execute($stmt);
mysqli_stmt_close($stmt);
mysqli_close($link);
```

Fig. 6 Code Snippet2 Backend

5.5 Linear Regression Technique

In statistics, linear regression is a linear approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X. The case of one explanatory variable is called simple linear regression(Refer Figure 7). For more than one explanatory variable, the process is called multiple linear regression.

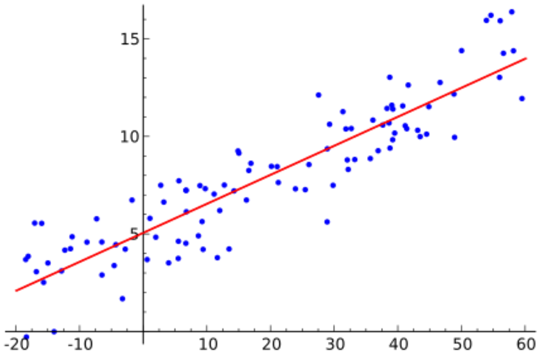
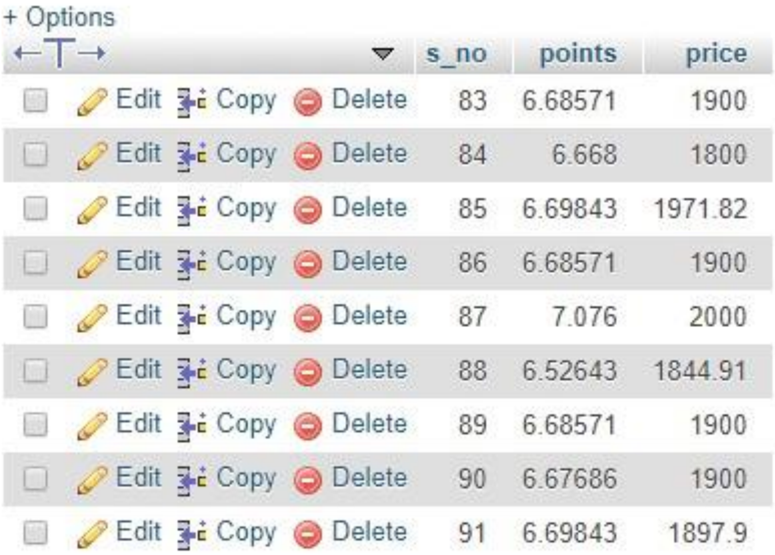


Fig. 7 Linear Regression Diag

```
function linear_regression( $x, $y ) {
 // linear regression function used in this model
   $n     = count($x);     // number of items in the array
   $x_sum = array_sum($x); // sum of all X values
   $y_sum = array_sum($y); // sum of all Y values
   $xx_sum = 0;
$xy_sum = 0;
   for($i = 0; $i < $n; $i++) {
      $xy_sum += ( $x[$i]*$y[$i] );
$xx_sum += ( $x[$i]*$x[$i] );
   }
   // Slope
   $slope = ( ( $n * $xy_sum ) - ( $x_sum * $y_sum ) ) / ( ( $n * $xx_sum ) - ( $x_sum *
$x_sum ) );     // calculate intercept
```

```
   $intercept = ( $y_sum - ( $slope * $x_sum ) ) /
$n;     return array(
        'slope'     => $slope,
        'intercept' => $intercept,    );}
```

Figure 8 shows the database table in the sense data is stored in the system with weight points as well as the prices.



+ Options

←T→	▼	s_no	points	price
□	Edit Copy Delete	83	6.68571	1900
□	Edit Copy Delete	84	6.668	1800
□	Edit Copy Delete	85	6.69843	1971.82
□	Edit Copy Delete	86	6.68571	1900
□	Edit Copy Delete	87	7.076	2000
□	Edit Copy Delete	88	6.52643	1844.91
□	Edit Copy Delete	89	6.68571	1900
□	Edit Copy Delete	90	6.67686	1900
□	Edit Copy Delete	91	6.69843	1897.9

Fig. 8 SQL Table

6. CONCLUSION

The proposed and developed model is capable enough to calculate property price for a single region property. The model is also flexible for all the regions.

Advantages

- **Every region has some places on which the residential prices are depending on. This model targets those places that too based on the user experiences of that place.**
- **Indirectly this model works like human approach. It first checks all the attributes which affect property prices and then predicts the prices, exactly like how a human mind works.**
- **This model is very flexible and Attributes can be changed, as there is always new construction in the town, which may or may not alter the prices. With this model those Attributes can be included and that too with the priority points.**
- **When this prediction model is used again and again for a region, it automatically adapts to that region and becomes more accurate with every use.**

In this manner, prices can be calculated in different regions and then can be divided into clusters on the basis of varying property prices. Hence, a universal modelling cluster can be obtained from the same approach.

6.1 Future Expansion

This work can be expanded by adding online Map API (Ex. Google API), in which we only need to select the location with one click and then the distance of that property from all the attributes can be calculated automatically with the help of API. There won’t be any need to go and actually survey the distances, which will save a lot of time and effort.

REFERENCES

[1] **Nissan Pow, Emil Janulewicz, Liu (Dave) Liu,** ***Spring 2017*****,** ***"Applied Machine Learning Project 4 Prediction of real estate property prices in Montreal", SJSU Scholar Works.***
http://rl.cs.mcgill.ca/comp598/fall2014/comp598_submission_99.pdf

[2] **Nihar Bhagat, Ankit Mohokar, Shreyansh Mane,** ***October 2016, "House Price Forecasting Using Data Mining", International Journal of Computer Applications Volume 152(2)***

[3] **Aaron Ng,** ***2015, "Machine Learning for a London housing Price Prediction Mobile Application", Thesis, Imperial College of London***