

# Prediction of Housing Prices under the Effect of Air Pollution

Ayushi Singh<sup>1</sup>, Divyanshu Garg<sup>2</sup>, Gaurav Multani<sup>3</sup>, Nipun Jain<sup>4</sup> and Anjali Sharma<sup>5</sup>

<sup>1-4</sup>Student, B. Tech Computer Science and Engineering, MIET College, Uttar Pradesh, India

<sup>5</sup>Assistant Professor, Dept. of Computer Science and Engineering, MIET College, Uttar Pradesh, India

\*\*\*

**Abstract** - In today's world, people are more conscious about their health due to the degradation of global environments such as air, water, and soil pollution. As the Air Quality Index (AQI) increases, the large percentage of the population is likely to experience increasingly severe adverse health effects. So that humankind can live without any dangerous conditions like poor AQI. In this paper, we explore the interaction between housing prices and air quality for 4 metropolitan cities using multiple regression. This report represents a website where considering the AQI we can choose our sweet livelihood house. We found evidence that good air quality is rewarded with higher housing prices and faster house price growth, in turn, it contributes to further air quality improvements. As we all know nowadays every individual wants to breathe clean air, and by using this website we can predict housing pricing on the basis of AQI and suggest people buy their house to their needs. In this paper, we will discuss the motivations and principles of various machine learning algorithms. These algorithms can be used for various purposes like prediction analyses, image processing, data mining, find patterns in data, etc. In this, we will review relevant meaning and present a conceptual framework and literature which elucidate the role of machine learning. Also, we will review and understand different types of supervised machine

**Key Words:** Air Quality Index, Air Quality, Housing Price, Machine Learning, Multiple Regression.

## 1. INTRODUCTION

Buying a house is an important decision of an individual's life as it requires a large amount of money and it also decides the standard of living conditions of a person, so it isn't important to make wiser choices that people are well aware of. As the society is advancing to a modern world people are getting more conscious to improve their living conditions. People are getting more conscious of environmental conditions like air pollution, greenery, sanitization, amenities, etc. near their living places. The need of improving the living conditions of the people and more awareness towards making better choices towards buying houses with advanced technologies like machine learning, predictive modeling available in today's world brought us to work on this project.

We can use machine learning to teach machines so that they can handle and process the data efficiently. We apply machine learning when we cannot interpret the pattern or extract or process the information in the given data just by viewing it. In this case, machine learning is used. With

plenty of datasets available, the demand for the technology of machine learning is in the rise. Many industries from the health industry to the military apply machine learning to extract relevant information. The main purpose of machine learning is to learn from the data and provides systems the ability to automatically learn and improve from experience without being explicitly programmed. There are many studies have been done on how to make machines learn by themselves. There are two techniques used in the field of machine learning, supervised learning, which trains a model on known input and output data so that it can predict future outputs, and unsupervised learning, which finds intrinsic structures in input data.

## 2. METHODOLOGY

Using Machine learning, the system have the ability to learn and improve from past experience automatically without being programmed in clear and detailed manner. Machine learning focuses on the development of computer programs that can access data and use it to learn from themselves. Multiple Regression is one of the approach of machine learning which is used in this project. It works on a number so the string inserted as a data is converted into numbers for prediction. Normalization of data is not necessary but it speeds up the calculation and sometime it may fail.

Equation of multiple linear regression:

$$y=A+BX_1+CX_2+DX_3$$

More than one independent variable is used to predict the dependent variable in multiple regression. Housing price is predicted by multiple regression on many factors.

## 3. LINEAR REGRESSION VS MULTIPLE REGRESSION(WITH EXAMPLE)

Linear regression is also called simple linear regression. It establishes the relationship between two different variables using a straight line. Linear regression try to draw a line that is closest to the data by finding the slope and intercept that minimize regression errors. In simple linear regression, the value of one or more independent variable is predicted using a single dependent variable.

$$\text{Equation : } Y=A+BX$$

Multiple regression can be explained when two or more independent variables are used to predict the value of a

dependent variable. The difference between linear regression and multiple linear regression is the number of independent variables.

Equation :  $y=A+BX_1+CX_2+DX_3$

Example- If an analyst wishes to establish a linear relationship between the daily change in a company's stock prices and other explanatory variables such as the daily change in trading volume or market returns. If he runs a regression analysis where the daily change in the company's stock prices is dependent variable and the daily change in trading volume is an independent variable, then this would be an example of a simple linear regression with one explanatory variable.

Whereas if the analyst adds the daily change in market returns as an independent variable into the regression, it would be a multiple linear regression.

4. PROPOSED SYSTEM

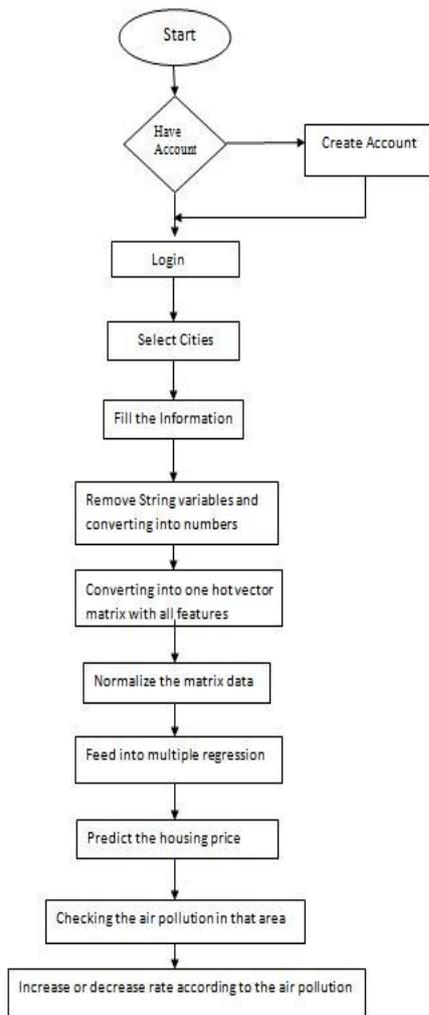


Fig. 1- Flow chart

5. INTERMEDIATE RESULT

The equation(Regression equation) that gives the relationships between a one dependent variable and one or more independent variables is given below:

$Y = A_0x_0 + A_1x_1 + A_2x_2 + \dots + A_kx_k$

In this equation, Y is the predicted value of the dependent variable. Values of the k independent variables are denoted by x1, x2, x3, ..., xk.

And finally, we have the A's - A0, A1, A2, ..., Ak. The A's are constants, called regression coefficients. Values are assigned to the A's based on the principle of least squares.

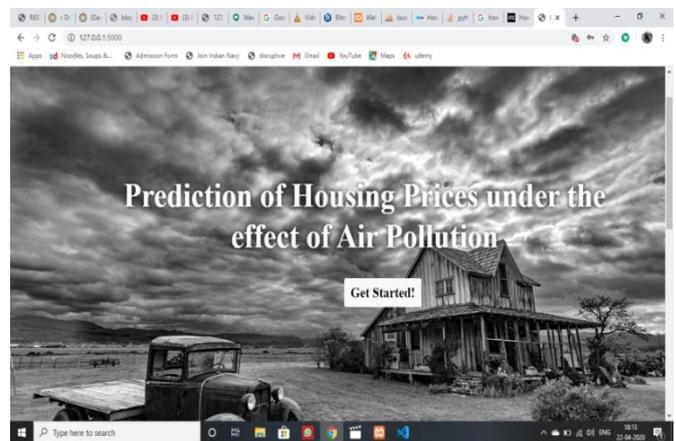


Fig-2: Homepage

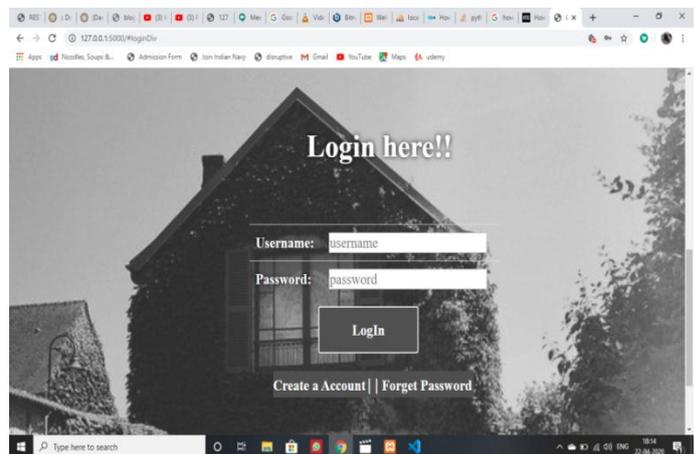


Fig-3: Login Page

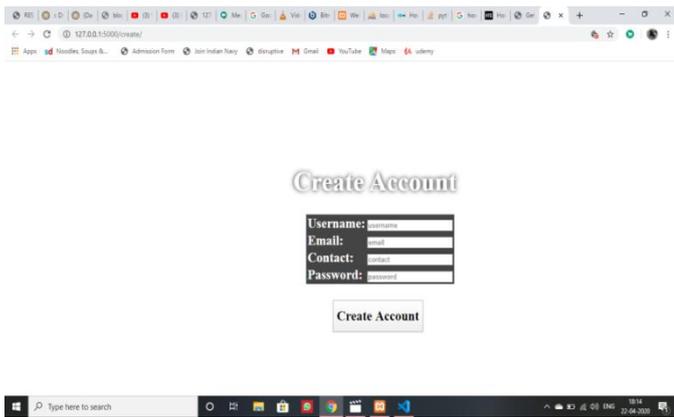


Fig-4: Creating new account



Fig-7: Information required to enter by user

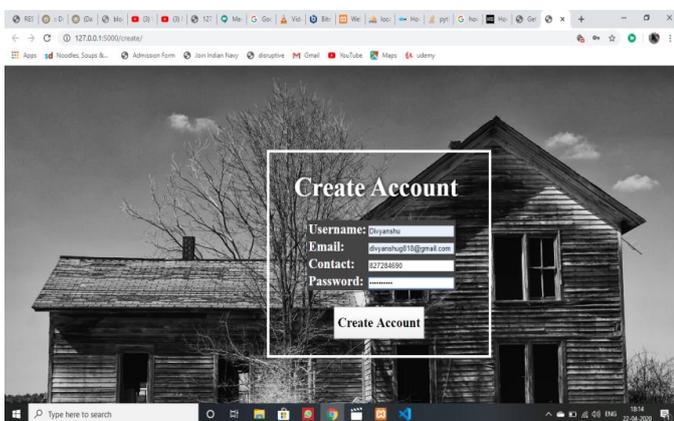


Fig-5: Create account page

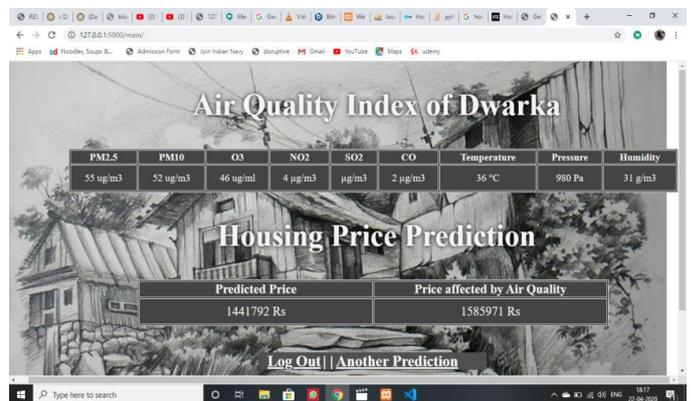


Fig-8: Results after prediction

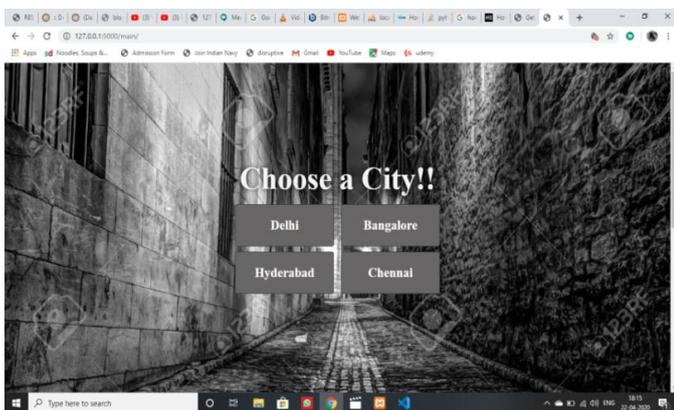


Fig-6: After entering details

### 6. EXPERIMENTAL RESULT

The table shown below is the predicted prices of some location and which is affected by air quality index like this we can get result of other location that is present.

Location	Upparpally, Hyderabad Expressway	Velachery	Dwarka
PM 2.5	76µg/m3	22µg/m3	55µg/m3
PM10	-µg/m3	3µg/m3	52µg/m3
O <sub>3</sub>	13µg/m3	1µg/m3	46µg/m3
NO <sub>2</sub>	3µg/m3	1µg/m3	4µg/m3
SO <sub>3</sub>	µg/m3	µg/m3	µg/m3
CO	3µg/m3	33µg/m3	2µg/m3
Predicted Price	851968Rs	861968Rs	1441792Rs
Price Affected by AQI	937164Rs	947164Rs	1585971Rs

Table-1: Result after prediction

## 7. FUTURE SCOPE

1. This application can be extended to every state, while it is limited to only a few metropolitan cities for now.
2. It can be deployed to a live server so that people can start using it at the mass level.
3. It tells about whether a person should buy a house at that location, in future a feature can be added that where should the person buy house other than the specified location

## 8. CONCLUSIONS

Housing prices predictor system is an application that provides a reliable platform to aware people of the actual price of housekeeping into consideration of their environmental choices. It predicts the position of a location which helps people to make to house buying decision wiser. The system will provide a general awareness among the people that living in an area having better environmental conditions is an important aspect of a person's life. It is a great platform for people having less knowledge about environmental and property prices to establish a link between the two. These steps are necessary to take to ensure sustainable development must be followed so as to have a better future.

Regression can be a really effective tool in statistical analysis of data. Different regression models can be used in different situations and their uses depends on a case to case, usually on the basis of type of data and the relationship between data. We can choose supervised learning if we need to train a model to make a prediction, for example, the future value of any variable, such as temperature, or a classification, for example, assigning different labels for different data items. And we can choose unsupervised learning if we need to explore our data and want to train a model to find a good representation, such as splitting data up into clusters (clustering). Also, Multiple linear regression is the most common form of linear regression. In prediction analysis, the multiple regression is the relationship between a dependent variable and two or more independent variables. These independent variables can be continuous or categorical (dummy coded as appropriate). And simple linear regression is used to predict the value of a dependent variable. The difference between the two is the number of independent variables.

## REFERENCES

[1] Ali Mohammad Khorshiddoust Predicted prices and environmental classification of economic values of selected areas of tabriz, International journal of economics and finance studies, vol 5, no 2, 2013 ISSN:1309-8055.

[2] Anselin L, Le Gallo J (2006) Interpolation of air quality measures in hedonic house price models: spatial aspects. *Spat Econ Anal* 1(1):31-52

[3] <https://stattrek.com/multiple-regression/regression-equation.aspx>

[4] [www.investopedia.com](http://www.investopedia.com)

[5] [www.aqicn.org](http://www.aqicn.org)

[6] Brasington DM, Hite D (2005) Demand for environmental quality: a spatial hedonic analysis. *Reg Sci Urban Eco* 35(1):57-82.

[7] Mark J, Goldberg M (1988) Multiple regression analysis and mass assessment: a review of the issues. *Apprais J* 56(1):89-109

[8] (Volume24) DOI: 10.4028/www.scientific.net/JERA.24.124 International Journal of Engineering Research in Africa

[9] A Review of Machine Learning and Deep Learning Applications, DOI: 10.1109/ICCUBEA.2018.8697857

[10] AcE-Bs 2012 Bangkok ASEAN Conference on Environment-Behaviour Studies, 16-18 July 2012

[11] [http://onlinestatbook.com/2/regression/multiple\\_regression.html](http://onlinestatbook.com/2/regression/multiple_regression.html)

[12] <https://www.mathworks.com/discovery/machine-learning.html>

[13] <https://medium.com/datadriveninvestor/regression-in-machine-learning-296caae933ec>

[14] [https://dss.princeton.edu/online\\_help/analysis/regression\\_intro.htm](https://dss.princeton.edu/online_help/analysis/regression_intro.htm)

[15] <https://blog.minitab.com/blog/how-to-choose-the-best-regression-model>

[16] <https://intellipaat.com/blog/tutorial/machine-learning-tutorial/classification-machine-learning/>

[17] <https://www.listendata.com/2018/03/regression-analysis.html#How-to-choose-the-correct-regression-model->