

NAME OF THE PROJECT

**HOUSING: PRICE PREDICTION**

Submitted by:

SAURABH SHRIVASTAVA

**ACKNOWLEDGMENT**

* To conduct this project the following tools have been used :
* Python 2.7
* Pandas (Library) : <http://pandas.pydata.org/>
* Numpy (Library) : <http://www.numpy.org/>
* Scikit­learn (Library) : [http://www.scikit­learn.org/](http://www.scikitlearn.org/)
* The Machine Learning part has been greatly inspired by the Machine Learning course teached by dr.Deepika Sharma of data trained (https://learning.datatrained.com/course/ml)
* Follow YouTube videos of Coursera.
* Follow Google different- different links.

**INTRODUCTION**

* Business Problem Framing
* Get the more effective feature that are effective on any prospective properties price so that company decide whether to invest in them or not.
* Features that is help to understand the price or value of any prospective

Houses are one of the necessary need of each and every person around the globe and therefore housing and real estate market is one of the markets which is one of the major contributors in the world’s economy. It is a very large market and there are various companies working in the domain.

.

Conceptual Background of the Domain Problem

* + Basic knowledge of real estate business.
  + Factors Affecting Housing Supply and Demand.
  + Understand real estate supply and demand of house.

* Review of Literature

This is a comprehensive summary of the research done on the topic. The review should enumerate, describe, summarize, evaluate and clarify the research done.

* Motivation for the Problem Undertaken

Being extremely interested in everything having a relation with the Machine Learning, the independent project was a great occasion to give me the time to learn and confirm my interest for this field. The fact that we can make estimations, predictions and give the ability for machines to learn by themselves is both powerful and limitless in term of application possibilities

This will be important for professionals who have not worked with huge dataset.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

Data Analysis

Feature Engineering

Feature Selection

Model Building

Model Deployment

* Data Sources and their formats
* Missing Values
* All The Numerical Variables
* Distribution of the Numerical Variables
* Categorical Variables
* Cardinality of Categorical Variables
* Outliers
* Relationship between independent and dependent feature

.

* Data Preprocessing Done

What were the steps followed for the cleaning of the data? What were the assumptions done and what were the next actions steps over that?

* Data Inputs- Logic- Output Relationships

Describe the relationship behind the data input, its format, the logic in between and the output. Describe how the input affects the output.

* State the set of assumptions (if any) related to the problem under consideration

Here, you can describe any presumptions taken by you.

* Hardware and Software Requirements and Tools Used

Listing down the hardware and software requirements along with the tools, libraries and packages used. Describe all the software tools used along with a detailed description of tasks done with those tools.

**CONCLUSION**

* Key Findings and Conclusions of the Study

We have completed project  successfully with **90%** accuracy using SVM which is great for the **‘House Price Prediction’** project. Now, we are ready to deploy our ML model in the real estate domain

* Learning Outcomes of the Study in respect of Data Science

### 

* Missing Value
* All The Numerical Variable
* Distribution of the Numerical Variable
* Categorical Variable
* Outliers
* Relationship between independent and dependent feature

SVM algorithms best

* The choice of algorithm depends on consideration of a number of factors such as the size of the data set, computing power of the equipment, and the availability of waiting time for the results. We recommend property valuers and researchers to use SVM for making forecasts if speed is a primary concern.

* Limitations of this work and Scope for Future Work

For future work, we trained all top supervised regression algorithms but we can try out a few of them which are always popular. After training all algorithms, we found that SVR have given high accuracy than remain but we have chosen SVR.