



AI PRODUCT/SERVICE PROTOTYPING FOR
SMALL BUSINESSES. (SOLO PROJECT)

[Redacted]

House Price predictions for Indian Real State Business

[Redacted]

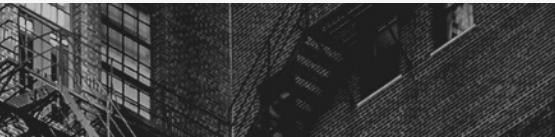
PREPARED BY
Abhinav Kumar Singh

Problem Statement

In India, there are multiple real estate small scale and large scale businesses that sell/buy/rent properties. These businesses spend a lot of time analyzing the prices of the property based upon its locality, services, and space. Sometimes this manual analysis leads to accurate price predictions and leads to the loss for the business.

This problem can be solved using Machine learning algorithms. Our goal is to solve this regression problem where the target variable is the price and the independent variables are number of bedrooms, number of bathrooms, super built-up area, carpet area, condition of house, area of population, locality, etc.





Business need assessment

There are more than 82,000+ real estate Companies in India which contributes 13% to the country's GDP. Automating this process not only saves time but will also benefits this business to conveniently sell/buy properties at correct prices and expand their business.



Target Specification

The target market for this product are the real state agents/ business/ portals which sell/buy properties. This product will use the previous data they have for the particular region and will use this to predict the buying/selling price of the new houses in the same region or having similar characteristics. dataset will contain:

Independent features

- no. of bedrooms
- no. of bathrooms
- condition/age of the house
- locality
- Land area
- Services provided
- parking availability
- prime location etc

Target feature

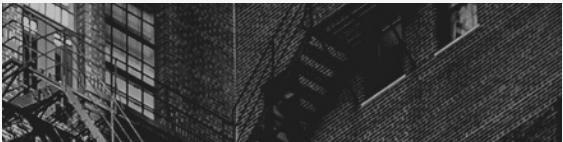
- Price of the house

External Search

The sources I have used as a reference for analyzing the need for such a system for local businesses and how real state business can use this technique to expand their business are:

- how Indian real estate works
<https://www.ibef.org/industry/real-estate-india.aspx>
- what is house price prediction and how to implement it
<https://towardsdatascience.com/create-a-model-to-predict-house-prices-using-python-d34fe8fad88f>
- Simple Dataset for implementation
<https://www.kaggle.com/alphaepsilon/housing-prices-dataset?select=train.csv>





Applicable Patents

Softwares we are going to use are

- Python
- Jupyter Notebook
- Tableau

Libraries used are

- NumPy
- Matplotlib
- Seaborn
- sklearn



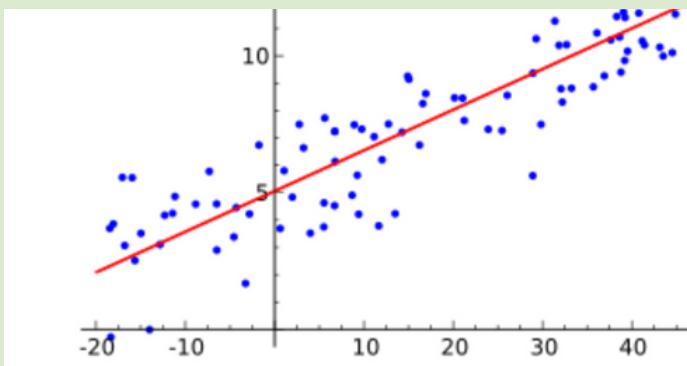
Application Constraints

- The model is not generalized. ie, if the dataset we use is for Bangalore houses, the trained model cannot be used to predict house prices for Mumbai or any other region.
- It will involve complex NLP if we take address as a feature to more accurately predict prices
- Need of correctly organised data of previously sold / bought houses which may contain some manual errors.

Algorithms Used

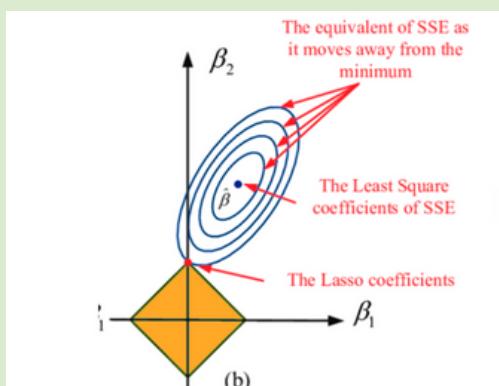
• Linear Regression

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting.



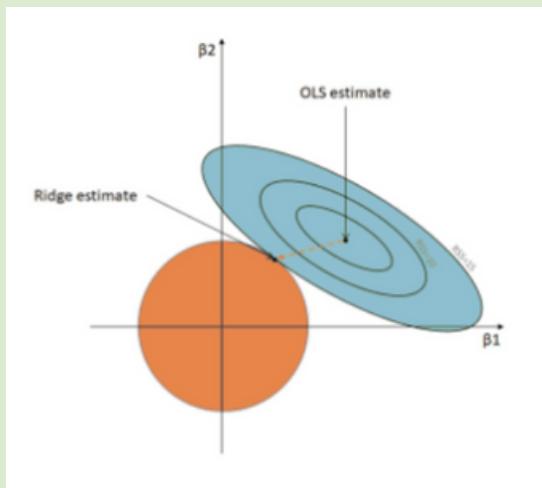
• Ridge and Lasso Regression

The word “LASSO” denotes Least Absolute Shrinkage and Selection Operator. Lasso regression follows the regularization technique to create prediction. It is given more priority over the other regression methods because it gives an accurate prediction.



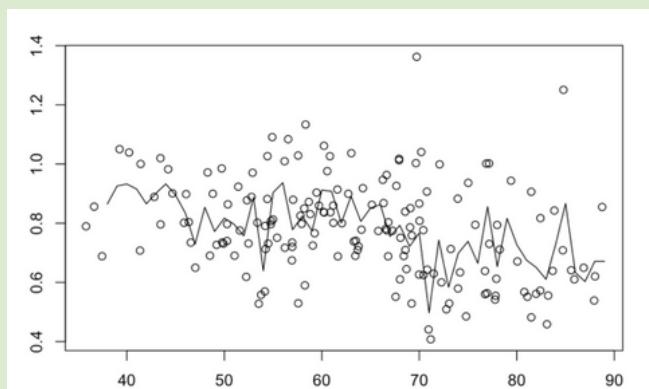
• Ridge and Lasso Regression

Ridge Regression is another type of regression algorithm in data science and is usually considered when there is a high correlation between the independent variables or model parameters. As the value of correlation increases the least square estimates evaluates unbiased values.



• K nearest Neighbors Regression

KNN regression is a non-parametric method that, in an intuitive manner, approximates the association between independent variables and the continuous outcome by averaging the observations in the same *neighborhood*.



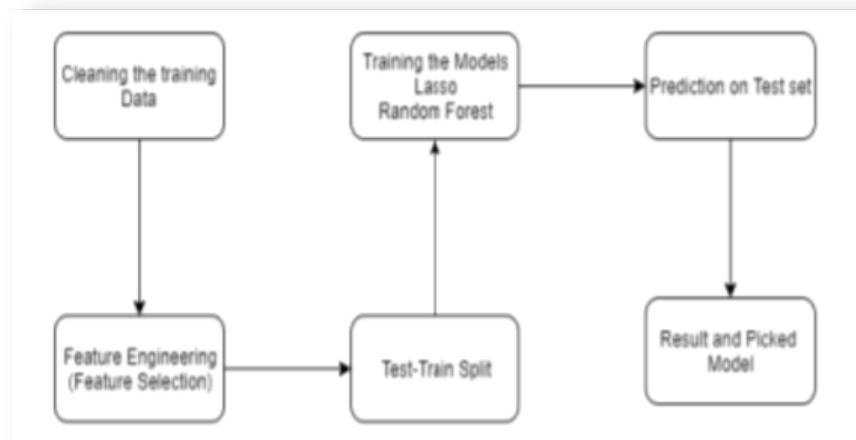
Final Prototype

Product Architecture

Simple flow



Detailed Flow



Code Implementation

This section include the code implementation for house price prediction problem on the kaggle housing dataset. In this First, I discussed the dataset and then Exploratory data analysis on the data , then i attached the code pdf file with this report.

About dataset

Dataset used is a kaggle housing price dataset which contains 80 dependent features and one dependent feature Saleprice. It consists of train.csv file for training the model and test.csv file for testing the model.

Some of the features are:

- YrSold
- SaleType
- SaleCondition
- Utilities
- LotArea
- Pool Area etc.

LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	PoolArea	PoolQC	Fence	MiscFeature	MiscVal	MoSold	YrSold	SaleType	SaleCondition	SalePrice
65.0	8450	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	2	2008	WD	Normal	208500
80.0	9600	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	5	2007	WD	Normal	181500
68.0	11250	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	9	2008	WD	Normal	223500
60.0	9550	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	2	2006	WD	Abnorml	140000
84.0	14260	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	12	2008	WD	Normal	250000
85.0	14115	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	MnPrv	Shed	700	10	2009	WD	Normal	143000
75.0	10084	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	8	2007	WD	Normal	307000
NaN	10382	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	Shed	350	11	2009	WD	Normal	200000
51.0	6120	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	4	2008	WD	Abnorml	129900
50.0	7420	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	1	2008	WD	Normal	118000

[Click on the Link for Code](#)