

HOUSING PRICES PREDICTION

DEPARTMENT OF CSE(AI & ML) TEAM - 8

20WH1A6637 - Thopuri Harshitha

20WH1A6638 - Dil Nawaz Farida

20WH1A6640 - Nelli Sanjana

20WH1A6642 - Gandham Bhagya sri

April 1, 2023

PROBLEM STATEMENT

- We all have experienced a time when we have to look up for a new house to buy. But then the journey begins with a lot of frauds, negotiating deals, researching the local areas and so on.
- So to deal with this kind of issues, we will be preparing a Machine Learning Based model, trained on the House Price Prediction Dataset. For each id in the test set, you must predict the value for the target Median House Value.
- This can be used by the buyer as well as seller for negotiation which is purely based on the data without intervention of any third party services.

DATASET

	id	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude	MedHouseVal
0	0	2.3859	15.0	3.827160	1.112100	1280.0	2.486989	34.60	-120.12	0.980
1	1	3.7188	17.0	6.013373	1.054217	1504.0	3.813084	38.69	-121.22	0.946
2	2	4.7750	27.0	6.535604	1.103175	1061.0	2.464602	34.71	-120.45	1.576
3	3	2.4138	16.0	3.350203	0.965432	1255.0	2.089286	32.66	-117.09	1.336
4	4	3.7500	52.0	4.284404	1.069246	1793.0	1.604790	37.80	-122.41	4.500

Figure: DataSet

PYTHON PACKAGES AND LIBRARIES

- Numpy
- Pandas
- Matplotlib
- Seaborn
- Geopandas
- Scikit-learn
- Geopy
- Flask

- Support Vector Machine
- Linear Regression
- Decision Tree Regressor
- Ridge Regressor
- Lasso Regression
- Polynomial Regression
- Random Forest Regressor

Random Forest Regression

- ① It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.
- ② One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables, as in the case of regression, and categorical variables, as in the case of classification

Polynomial Regression

- 1 Linear regression requires the relation between the dependent variable and the independent variable to be linear. What if the distribution of the data was more complex?
- 2 Polynomial regression is a form of Linear regression where only due to the Non-linear relationship between dependent and independent variables, we add some polynomial terms to linear regression to convert it into Polynomial regression.

Ridge Regression

- 1 Ridge regression is a method of estimating the coefficients of multiple-regression models in scenarios where the independent variables are highly correlated.
- 2 Overall, ridge regression is a useful technique for datasets with multicollinearity and/or high-dimensional data, where it can help to improve the accuracy and generalization performance of the regression model.

Decision Tree Regression

- 1 Decision tree builds regression or classification models in the form of a tree structure.
- 2 It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.

OUTPUT

	id	MedHouseVal
0	37137	0.664100
1	37138	1.095900
2	37139	3.366000
3	37140	3.799905
4	37141	2.381100
...
24754	61891	1.993300
24755	61892	2.068300
24756	61893	1.047500
24757	61894	3.797902
24758	61895	3.226600

24759 rows × 2 columns

Comparison Graph

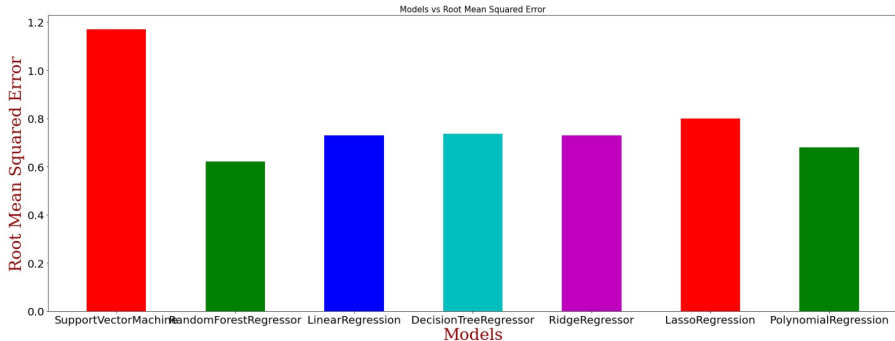
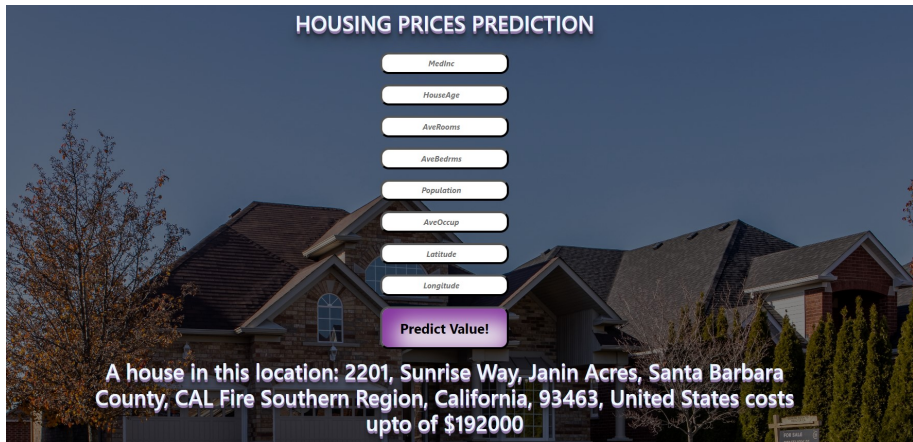


Figure: Models Comparison

Comparison Table

MODEL NAME	MSE	RMSE
Support Vector Machine	1.36	1.16
Random Forest Regression	0.38	0.61
Linear Regression	0.53	0.72
Decision Tree Regression	0.54	0.73
Ridge Regression	0.53	0.72
Lasso Regression	0.63	0.79
Polynomial Regression	0.46	0.68

The image shows a web application interface for housing price prediction. The background is a photograph of a large, two-story house with a dark roof and brick accents. Overlaid on this image is a form with the title "HOUSING PRICES PREDICTION" at the top. Below the title are eight input fields, each with a label: "MedInc", "HouseAge", "AveRooms", "AveBedrms", "Population", "AveOccup", "Latitude", and "Longitude". These labels are in a small, italicized font. Below the input fields is a purple button with the text "Predict Value!". At the bottom of the form, there is a text output area that reads: "A house in this location: 2201, Sunrise Way, Janin Acres, Santa Barbara County, CAL Fire Southern Region, California, 93463, United States costs upto of \$192000".

HOUSING PRICES PREDICTION

MedInc

HouseAge

AveRooms

AveBedrms

Population

AveOccup

Latitude

Longitude

Predict Value!

A house in this location: 2201, Sunrise Way, Janin Acres, Santa Barbara County, CAL Fire Southern Region, California, 93463, United States costs upto of \$192000

Website link: <http://thopuriharshitha.pythonanywhere.com>

THANK YOU