Vijayakumar kurup

This document details about the project which predicts the house price using various ML algorithm and way to compare them.

California House price prediction

Python version

## Table of Contents:

[Table of Contents: 1](#_Toc47090380)

[Introduction: 2](#_Toc47090381)

[Included with the project: 2](#_Toc47090382)

[Getting Started: 2](#_Toc47090383)

[Pre-requisistes: 2](#_Toc47090384)

[Installation: 2](#_Toc47090385)

[Usage: 3](#_Toc47090386)

[Input File: 3](#_Toc47090387)

[Implement the code : 3](#_Toc47090388)

[**Jupyter Notebook:** 3](#_Toc47090389)

[**Command Prompt and Python file:** 3](#_Toc47090390)

[Output 4](#_Toc47090391)

[**Local Webpage** 4](#_Toc47090392)

[Contact: 4](#_Toc47090393)

[Acknowledgements: 4](#_Toc47090394)

## Introduction:

"This Capstone project is created for predicting the house prices for any neighborhood. The dataset used for this project is a comma separated file called “cal\_housing with header.csv”. Several ML algorithms were used in doing this.

### Included with the project:

This project includes the input dataset, the Jupyter notebook, the necessary python packages, and a client implementation.

## Getting Started:

The data set appears to be bit confusing as it shows the number of rooms in 10s and 100s. Also, other details which is in the data set are much more than the normal house hold. All it means that it takes a neighborhood and provide the details about the sum of the items from each individual house.

## Pre-requisites:

To run this project and evaluating various algorithms, some basic computer knowledge along with knowledge of running a python project, and knowledge of prediction logics and HTML client application.

## Installation:

Install the following packages

Python 3.7

streamlit

pandas

joblib

sklearn

or use the requirement.txt for the installation if IDE like PyCharm is used.

## Usage:

From any command prompt like conda with environment where all the above packages given in the installation section, are installed and then run the following command

“streamlit run homepredictor.py”

### Input File:

There are two aspects of this project.

1. Create the prediction model.

To do this Jupyter notebooks are provided for the prediction logic. It takes the input data set and create the required model file called “XXX.pkl”. XXX denotes the output file related to the kind of model used for this purpose. It also creates the scaler file gives the details about how the inputs are scaled before fitting the models.

1. Use the prediction model

Using the model and scaler file (created by the Jupyter notebook) predict the house prices from various models used.

### Implement the code :

**Jupyter Notebook:**

1. Load the data from cal\_housing with header.csv

2. Preprocess the data by removing the missing values, null values.

3. Find correlation between the features for the selection of features.

4. split the data train vs test 70-30 ratio.

5. From here it takes multiple branches. ( Save the scaler information in respective YYY.save file YYY gives the kind of regularization used.)

A. Do Normalization

B. Do standardization

C. Without any regularization

6. Another deviation is to perform various modellign algorithms like Linear regression, Ridge Regression, Lasso Regression, Elastic net regression, Random Forest Regression, SVR and ADA Boost Regressions. Similar ones are grouped in one notebook.

7. Save the pickle file and the scaler files for the client to use for the predictions from user inputs.

8. The client application is created using streamlit and the requirement file with the details in the session Installation.

9. How to make it work::

Copy following files xx.pkl, scaler file xx.save, homepredictor.py.

eg. VK\_ELASTIC\_Model\_mm.pkl elastic = elasticnet, mm = min-max scaler CV = cross validation used etc.

Capstone California House-LR-RR-LASSO-NET - Zscore standardization.ipynb

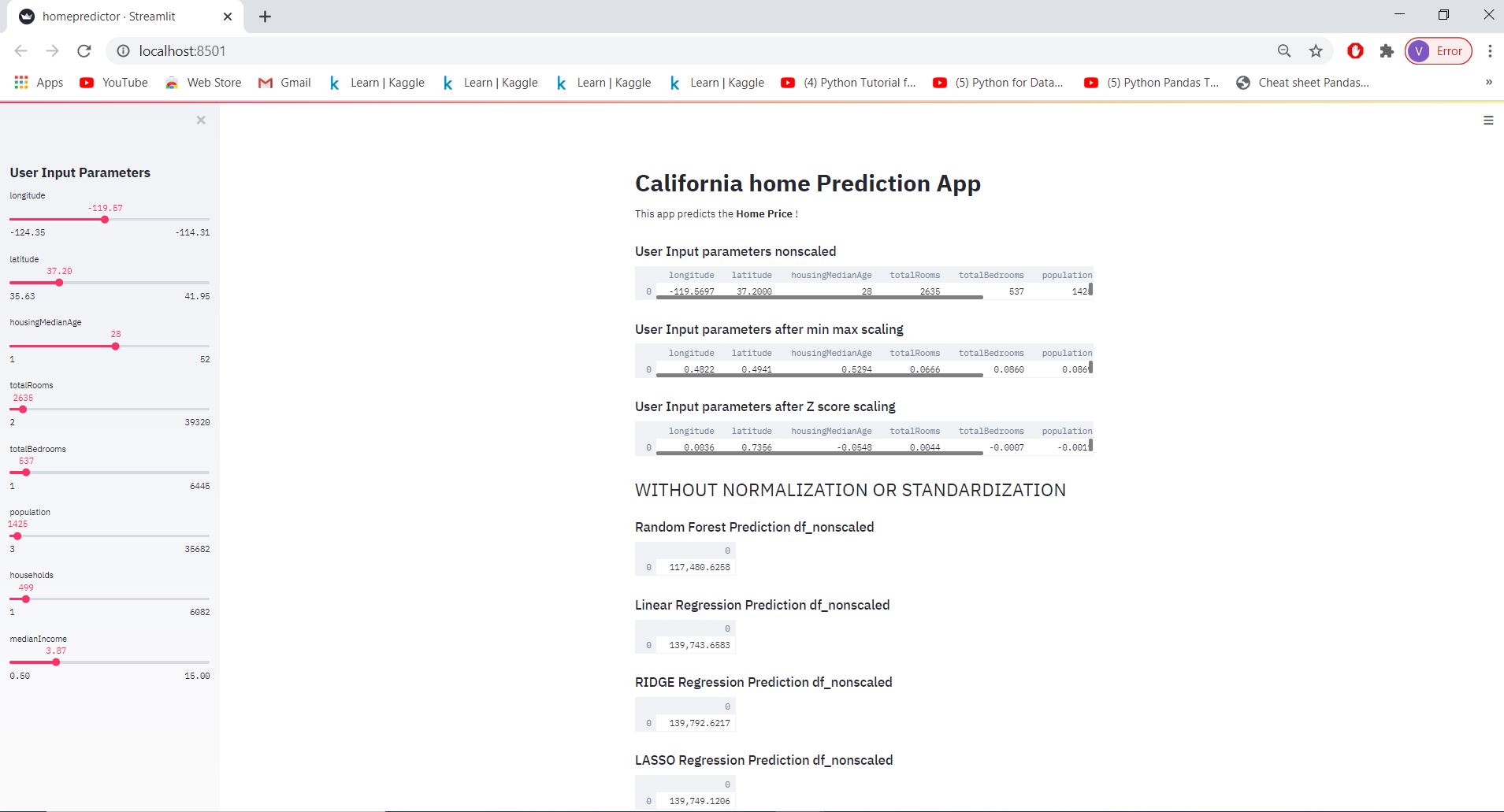
LR – Linear regression, RR = Ridge, Lasso = lasso regression, Zscre is the standardization etc.

**Command Prompt and Python file:**

Refer Usage section of this document how to use client application.

### Output

**HTML page using the local host. (A typical output)**

[](https://raw.githubusercontent.com/vkkurup/Project/master/Cal-state-house-price-predictor/sample%20output.JPG)

## Enhancements:

The client application will be upgraded for publishing the application on Heroku server website. In that case we don’t need to copy or download the files from the GitHub.

## Contact:

Any questions Please contact [Vijayakumar Kurup](mailto:vkkurup@gmail.com)