USA HOUSING DATASET

Some information about a bunch of houses in regions of the United States , it is all in the data set: USA_Housing.csv.

The data contains the following columns:

'Avg. Area Income': Avg. Income of residents of the city house is located in. 'Avg. Area House Age': Avg. Age of Houses in same city 'Avg. Area Number of Rooms': Avg. Number of Rooms for Houses in same city 'Avg. Area Number of Bedrooms': Avg. Number of Bedrooms for Houses in same city 'Area Population': Population of city house is located in 'Price': Price that the house sold at 'Address': Address for the house.

PROJECT OVERVIEW

In this project supervised learning techniques is applied on a dataset of houses, it's characteristics and pricing in USA. The dataset is explored by selecting small subset as sample. Then it is preprocessed y scaling each feature and identifying unwanted outliers. Now, the data is found cleaned and the features and outcome are trained and fitted. Predictions are visualized using graph plotting and coefficient of determination or r-squared value of the model is determined.

SOFTWARES AND LIBRARIES

- Python 3
- Pandas
- Seaborn
- Matplotlib
- Scikit-learn
- Jupyter notebook

AADARSH KUMAR SINGH 03 MAY 2020