Predicting house prices using machine learning

# Design thinking:

## Data source:

In the house pricing the datasets are there they are

* [**Pandas**](https://www.geeksforgeeks.org/python-pandas-dataframe/)**–** To load the Dataframe
* [**Matplotlib**](https://www.geeksforgeeks.org/matplotlib-tutorial/)**–** To visualize the data features i.e. barplot
* [**Seaborn**](https://www.geeksforgeeks.org/introduction-to-seaborn-python/)**–** To see the correlation between features using heatmap

## 2.DataPreprocessing:

 we categorize the features depending on their datatype (int, float, object) and then calculate the number of them.

## Exploratory Data Analysis

[EDA](https://www.geeksforgeeks.org/what-is-exploratory-data-analysis/) refers to the deep analysis of data so as to discover different patterns and spot anomalies

To analyze the different categorical features. By barplot

## Data Cleaning

[Data Cleaning](https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/) is the way to improvise the data or remove incorrect, corrupted or irrelevant data

|  |
| --- |
|  |

## 3.Feature collection:

## OneHotEncoder – For Label categorical features

One hot Encoding is the best way to convert categorical data into binary vectors.

## 4.Model Selection:

As we have to train the model to determine the continuous values, so we will be using these regression models.

* SVM-Support Vector Machine
* Random Forest Regressor
* Linear Regressor

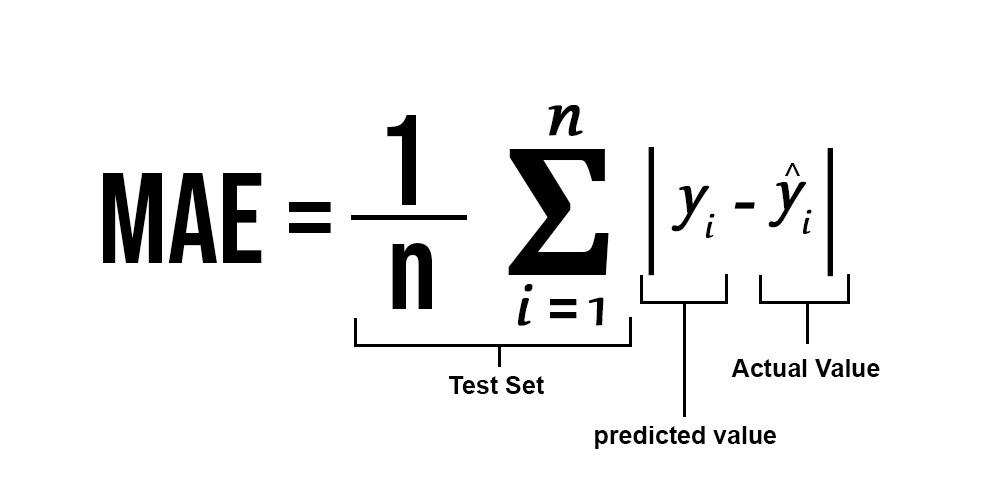
## And To calculate loss we will be using the [mean\_absolute\_percentage\_error](https://www.geeksforgeeks.org/how-to-calculate-mape-in-python/) module

## 5.Model Training:

## Splitting Dataset into Training and Testing

X and Y splitting (i.e. Y is the SalePrice column and the rest of the other columns are X)

## 6:Evaluation:



## Conclusion:

Clearly, SVM model is giving better accuracy as the mean absolute error is the least among all the other regressor models i.e. 0.18 approx. To get much better results ensemble learning techniques like [Bagging](https://www.geeksforgeeks.org/ml-bagging-classifier/) and [Boosting](https://www.geeksforgeeks.org/xgboost/) can also be used.