

# Regression code:

## Project implementation steps:

- 1) Download home price prediction dataset from <https://www.kaggle.com/>.
- 2) Import modules (pandas, sklearn, numpy, seaborn).
- 3) Reading dataset and visualizing it.
- 4) Preprocessing dataset:
  - Remove Missing values.
  - Convert object attributes to numerical attributes.
  - Remove Duplicated records.
  - Calculating Correlation attributes.
  - Remove attributes that big positive correlation and big negative correlation.
  - Saving clean dataset in new csv\_file.
- 5) Splitting dataset to training dataset and testing dataset.
- 6) Scaling training dataset and testing dataset to apply regression algorithms.
- 7) Apply linear **regression algorithms**.
- 8) Calculate accuracy: the result is: **63%** .
- 9) Draw a forecast chart **linear Regression**.
- 10) Apply **random forest regression** algorithms.
- 11) Calculate accuracy: the result is: **68%** .
- 12) Draw a forecast chart **random forest**.

## Results:

Random forest regressor work well on large dataset and it has a complex function, so that it gives good results compare with linear regression which has a simple equation.

## Project implementation steps:

- 1) Download cancer dataset from <https://www.kaggle.com/>.
- 2) Import modules (pandas, sklearn, numpy, seaborn).
- 3) Reading dataset and visualizing it.
- 4) Adding noise to dataset.
- 5) Preprocessing dataset:
  - Remove Missing values.
  - Remove Duplicated records.
  - Calculating Correlation attributes.
  - Remove attributes that big positive correlation and big negative correlation.
  - Saving clean dataset in new csv\_file.
- 6) Splitting dataset to training dataset and testing dataset.
- 7) Apply **K-Neighbors classifier** Algorithm.
- 8) Calculate accuracy: the result is: **90%** .
- 9) Calculate precision and recall and f1\_score for **K-Neighbors classifier**.
- 10) Draw a forecast chart **K-Neighbors classifier**.
- 11) Apply **Decision Tree Classifier** algorithm.
- 12) Calculate accuracy: the result is: **95%** .
- 13) Calculate precision and recall and f1\_score for **Decision Tree Classifier**.
- 14) Draw a forecast chart **Decision Tree Classifier**.

## Results:

Both algorithms work well but **Decision Tree Classifier** take the entire input training data root and Don't dependent on randomness so that it accuracy is higher than **K-Neighbors**.

