Kaggle Competition Link: <https://www.kaggle.com/c/house-prices-advanced-regression-techniques/overview>

Day 1: Learnt python

Day 2: Loan prediction sample project practise(Regression model)

Day 3: Iris data set practise (classification problem)

Day 4: **Joined in Kaggle.Downloaded data set for “House Prices Advanced Regression Techniques” competition.Started model building.**

Feature Engineering

* Set the relevant data types
* Handled missing values
* Handled outliers
* Changed categorical values into numerical values using Label Encoding technique.
* As Sales price values are right skewed the values needed to be more normally distributed. That was achieved using Log-Transformation technique.
* Analyses correlation between features

Day 5: The predictive model has been built. Used **Ensemble Learning Technique** such as

Stacking, Weighted average and Boosting advanced regression techniques.

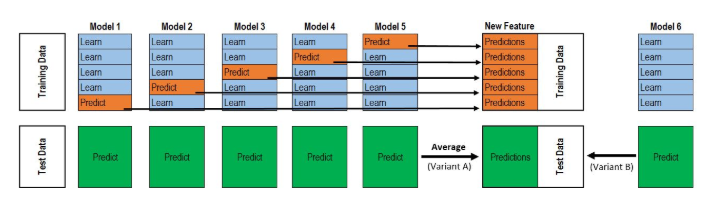
The base regression models that have been used are

* Lasso Regression
* ElasticNet(ENet) Regression
* Kernel Ridge Regression
* Gradient Boosting Regression
* XGBoost
* LightGBM

**Ensemble Learning**

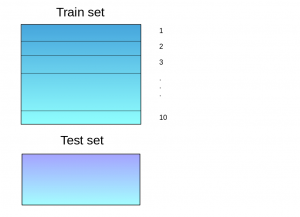
Diverse set of models in comparison to single models to make better decisions.This diversification in Machine Learning is achieved by Ensemble Learning.Various techniques in ensemble learning are given below.

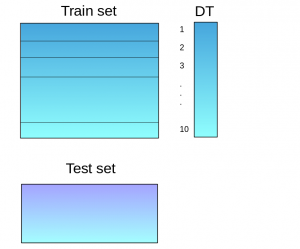
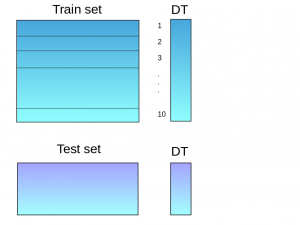
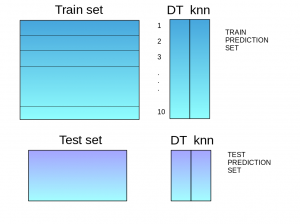
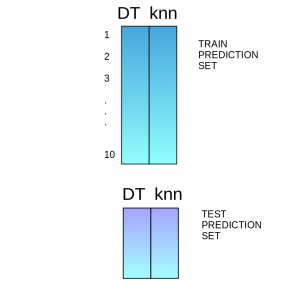
1. Max Voting
2. Averaging
3. Weighted Averaging
4. Stacking
5. Blending
6. Bagging
7. Boosting

**STACKING**

Stacking is an ensemble learning technique that uses predictions from multiple models (for example decision tree, knn or svm) to build a new model. This model is used for making predictions on the test set. Below is a step-wise explanation for a simple stacked ensemble:

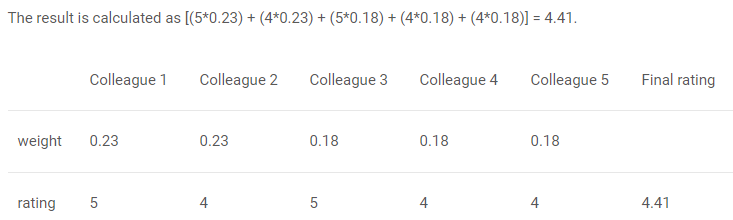
1. The train set is split into 10 parts.



1. A base model (suppose a decision tree) is fitted on 9 parts and predictions are made for the 10th part. This is done for each part of the train set.  
   
2. The base model (in this case, decision tree) is then fitted on the whole train dataset.
3. Using this model, predictions are made on the test set.  
   
4. Steps 2 to 4 are repeated for another base model (say knn) resulting in another set of predictions for the train set and test set.  
   
5. The predictions from the train set are used as features to build a new model.  
   
6. This model is used to make final predictions on the test prediction set.

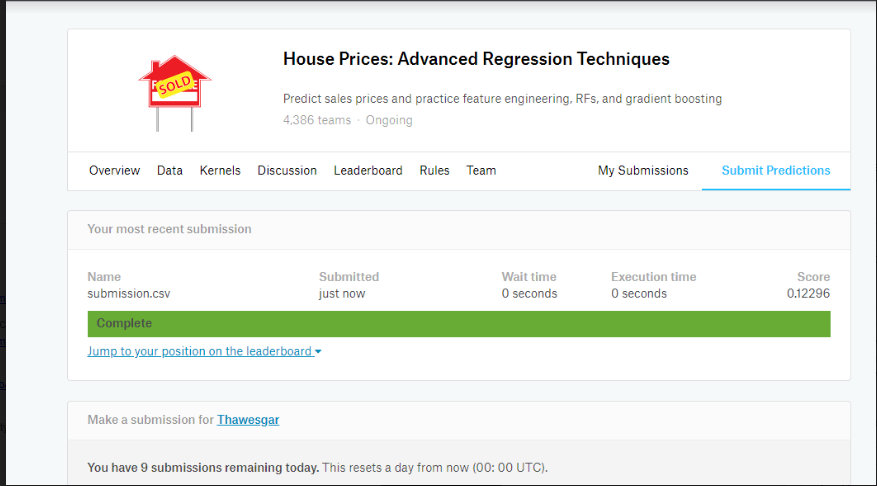
**WEIGHTED AVERAGING**

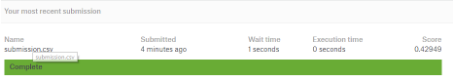
This is an extension of the averaging method. All models are assigned different weights defining the importance of each model for prediction.Eg:

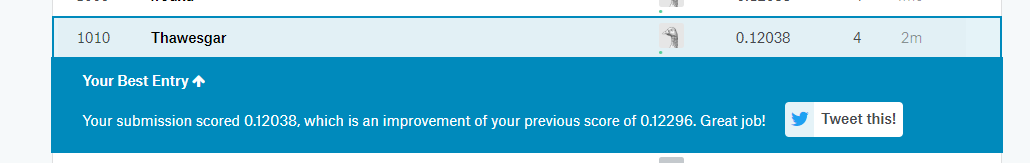
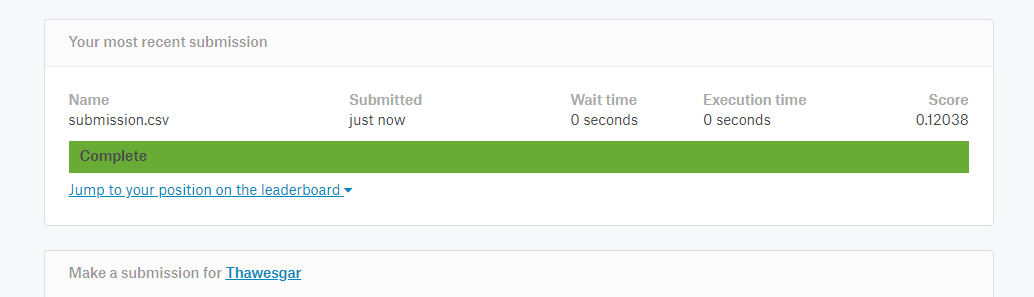


Day 6: Submitted the submission.csv file to kaggle and prepared the document to submit in course web.

**1st submission in kaggle and the score:**

**2nd Submission**

****

**Last submission in kaggle and the score**

**Current score out of hundred: ((1-0.12038=0.87962)\*100=87.962%)**

**Reference**

A Comprehensive Guide to Ensemble Learning (With Python codes) : <https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/>