# Assignment – Ensemble Models

1. What is an Ensemble Model?
2. What is the difference between Bagging and Boosting. Explain it using an example?
3. Explain the concept of Bias Variance Trade Off?
4. Try Ensemble & find the Best Model for **Motor Trend Car Dataset**. The Lowest MSE found is 7.9. (Tip: Keep CV = 20 and Random State = 123)

* mpg - Miles/(US) gallon – Target Variable
* cyl - Number of cylinders
* disp - Displacement (cu.in.)
* hp - Gross horsepower
* drat - Rear axle ratio
* wt - Weight (1000 lbs)
* qsec - 1/4 mile time
* vs - Engine (0 = V-shaped, 1 = straight)
* am - Transmission (0 = automatic, 1 = manual)
* gear - Number of forward gears
* carb – No of Carburetors

1. Download the dataset from Kaggle under the following link:

<https://www.kaggle.com/uciml/pima-indians-diabetes-database>

Apply Ensemble & find the Best Model. The Highest Accuracy Score is 0.7736

(Tip: Use Random State = 123 & Cross Validation as 20)

Data Dictionary – Diabetes Dataset

Predictor Variables:

* **Pregnancies** - Number of times pregnant
* **GlucosePlasma** - Glucose Concentration 2 hours in an oral glucose tolerance test
* **BloodPressure** - Diastolic blood pressure (mm Hg)
* **SkinThickness** - Triceps skin fold thickness (mm)
* **Insulin** - 2-Hour serum insulin (mu U/ml)
* **BMI** - Body mass index (weight in kg/(height in m)^2)
* **DiabetesPedigreeFunction** - Diabetes pedigree function
* **Age** - Age in (years)

Target Variable:

* **Outcome** - Class variable (0 or 1) 268 of 768 are 1, the others are 0

1. Apply Ensemble Modelling on Titanic Dataset. The data has been sent along. You need to apply the Ensembling and find the optimum model for the data. Find out the best model using HyperTuning of the Parameters followed by Grid Search.