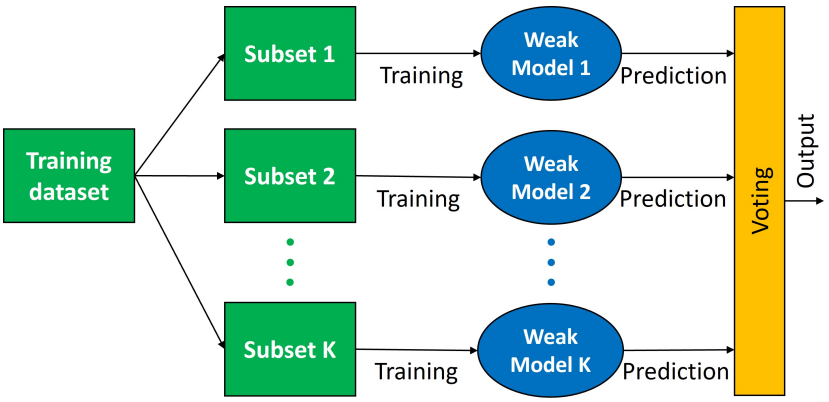
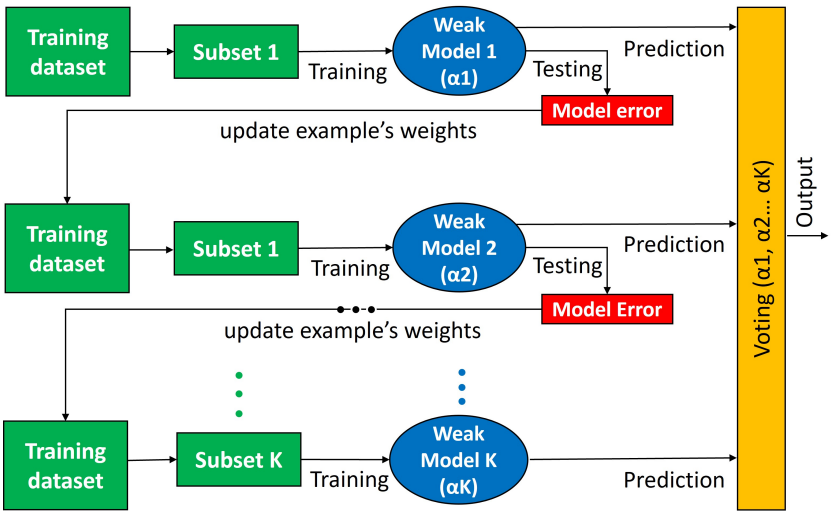
Ensemble Methods

Ensemble method is a technique that uses multiple machine learning models in order to obtain better performance. It includes reducing the variance, reducing the bias or increasing accuracy.

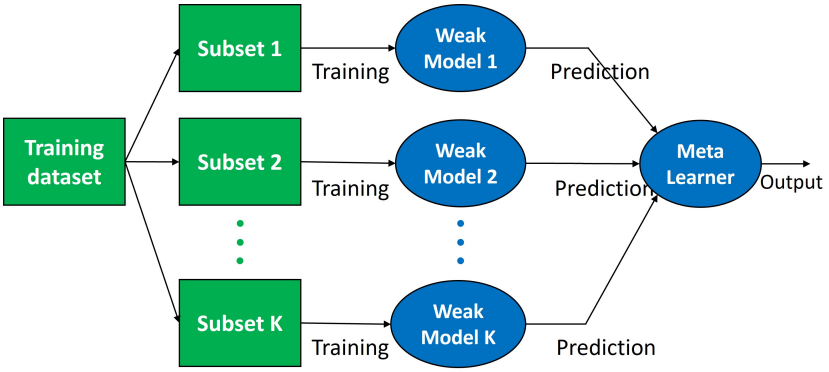
Bagging (Bootstrap aggregation):
Step 1: Create K subsets from the training dataset with random sampling. Each subset is sampled with the replacement of instances back. This is called row sampling with replacement.
Step 2: For each subset, train K weak models independently. These models are homogeneous, which means these models are of the same type.
Step 3: For inference, take the predictions from each model and aggregate them into a single prediction using averaging or max voting.





Boosting:
Step 1: Assign an equal weight to each example in training and create the first subset using row sampling.
Step 2: Using the first subset, Train the first weak model independently using focal loss. Update weights of training dataset examples and model alpha value using model error.
Step 3: Get another subset from the weights updated training dataset (so that false predicted examples are selected again) and perform step 2. Keep doing it from K homogenous models.
Step 4: For inference, take the predictions from each model and aggregate them into a single prediction using a weighted average.

Stacking:
Step 1: Create K subsets from the training dataset with row sampling with replacement.
Step 2: For each subset, train K weak models independently. These models are heterogenous, which means these models are of different types.
Step 3: Take the predictions from each model and create a new training dataset which will be used to train the meta-model. For inference, take prediction from weak models then meta-model.



Summary:

	Bagging	Boosting	Stacking
Used for	Reduce Variance	Reduce Bias	Increase accuracy
Weak Models Type	Homogenous	Homogenous	Heterogenous
Voting Type	Max, Averaging	Weighted Averaging	Meta-Model

For any mistake/query: alishafique3@gmail.com For references and updates: Website, Github