stackfuel

Ensembling:
Let’s work together

Focus Session. April 24th, 2024.

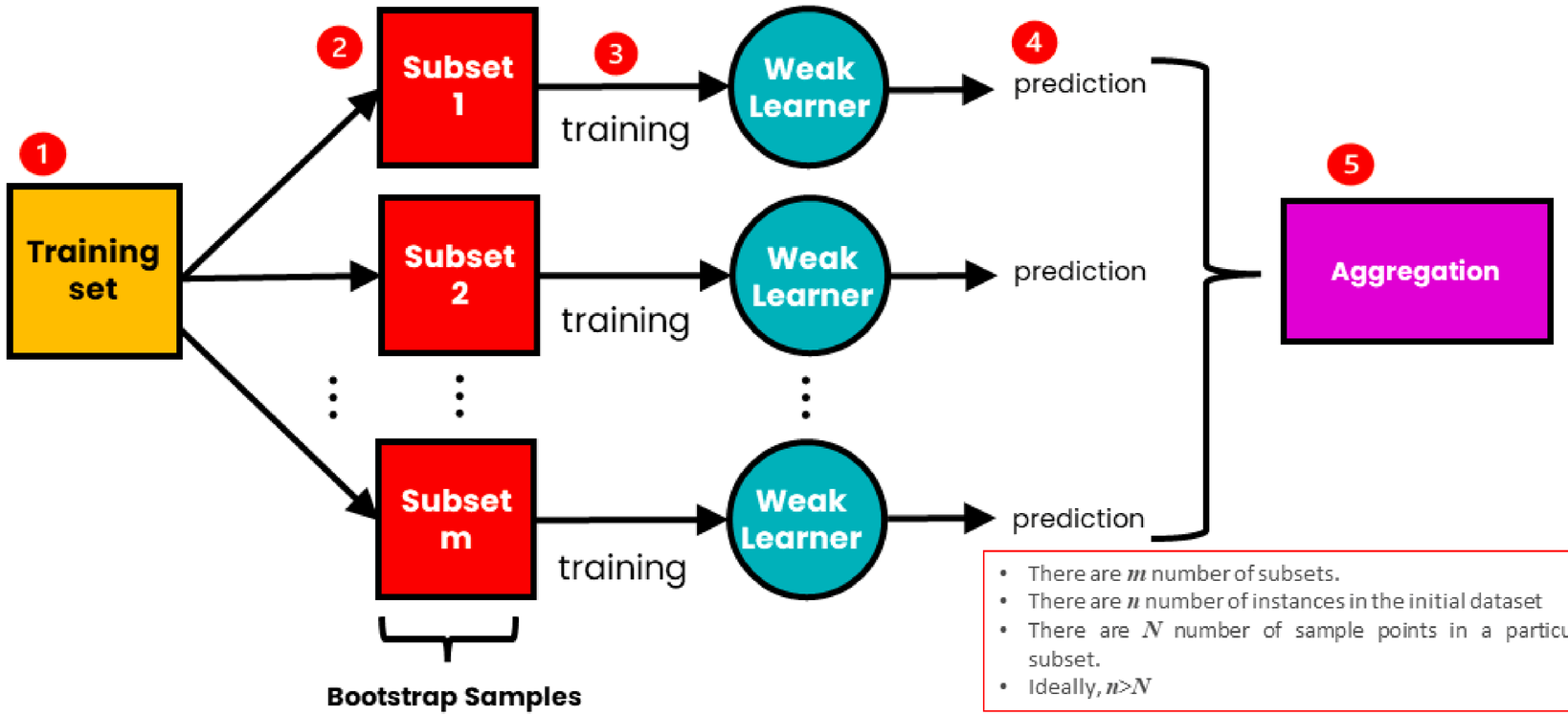
stackfuel

We use bagging for combining weak learners of high variance. Bagging aims to produce a model with lower variance than the individual weak models.

Ensemble Learning Methods

Bootstrap Aggregation (Bagging)

Models are homogeneous



- There are m number of subsets.
- There are n number of instances in the initial dataset
- There are N number of sample points in a particular subset.
- Ideally, $n>N$

Subsets are taken from initial dataset by resampling with replacement.

Each of these bootstrapped subsets is used to train a weak learner. Each model trains individually.

Resulting predictions are aggregated at the end using max(voting) or averaging.

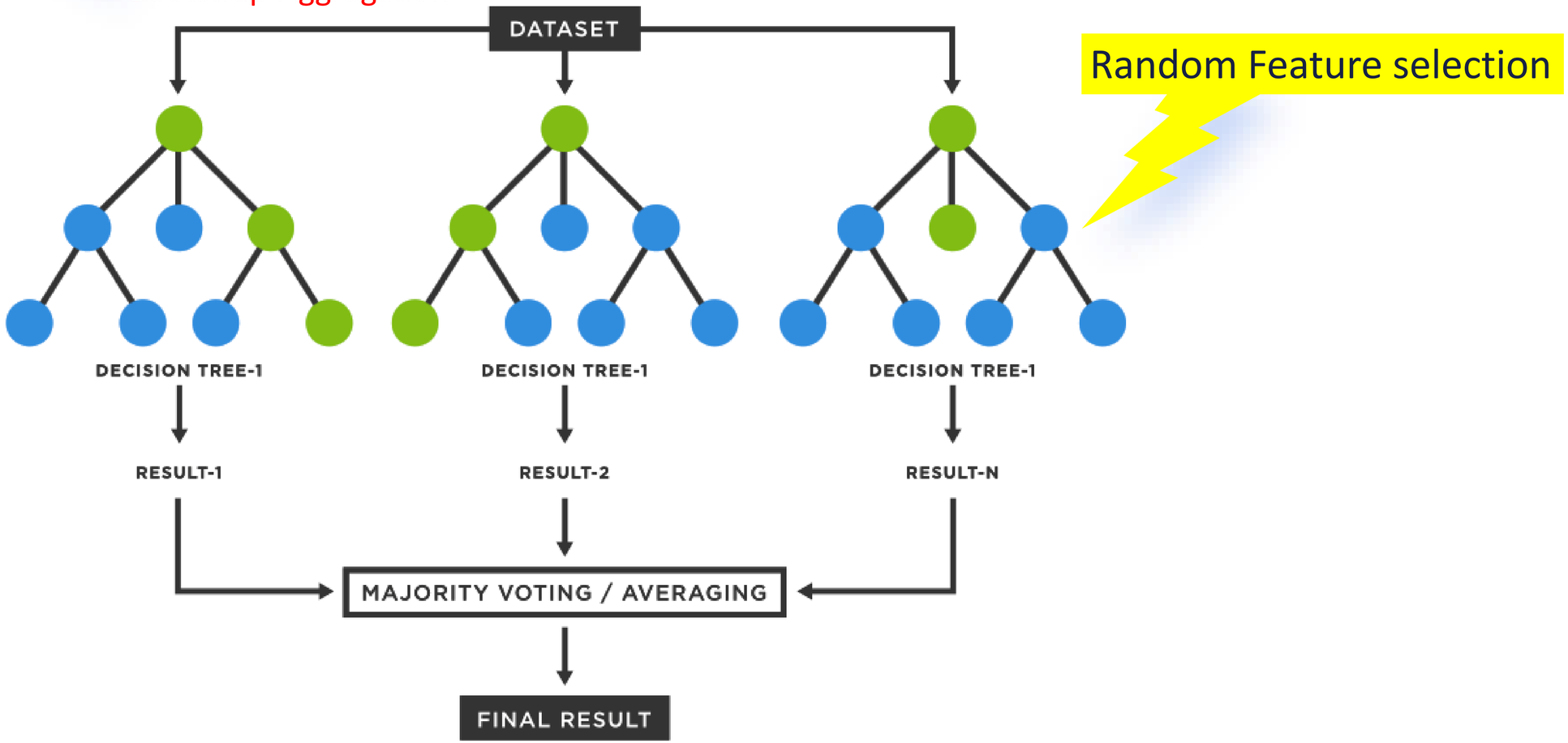
Ensemble Learning Methods

Random Forest

Random Sampling

Bootstrap-Aggregation

Random Feature selection

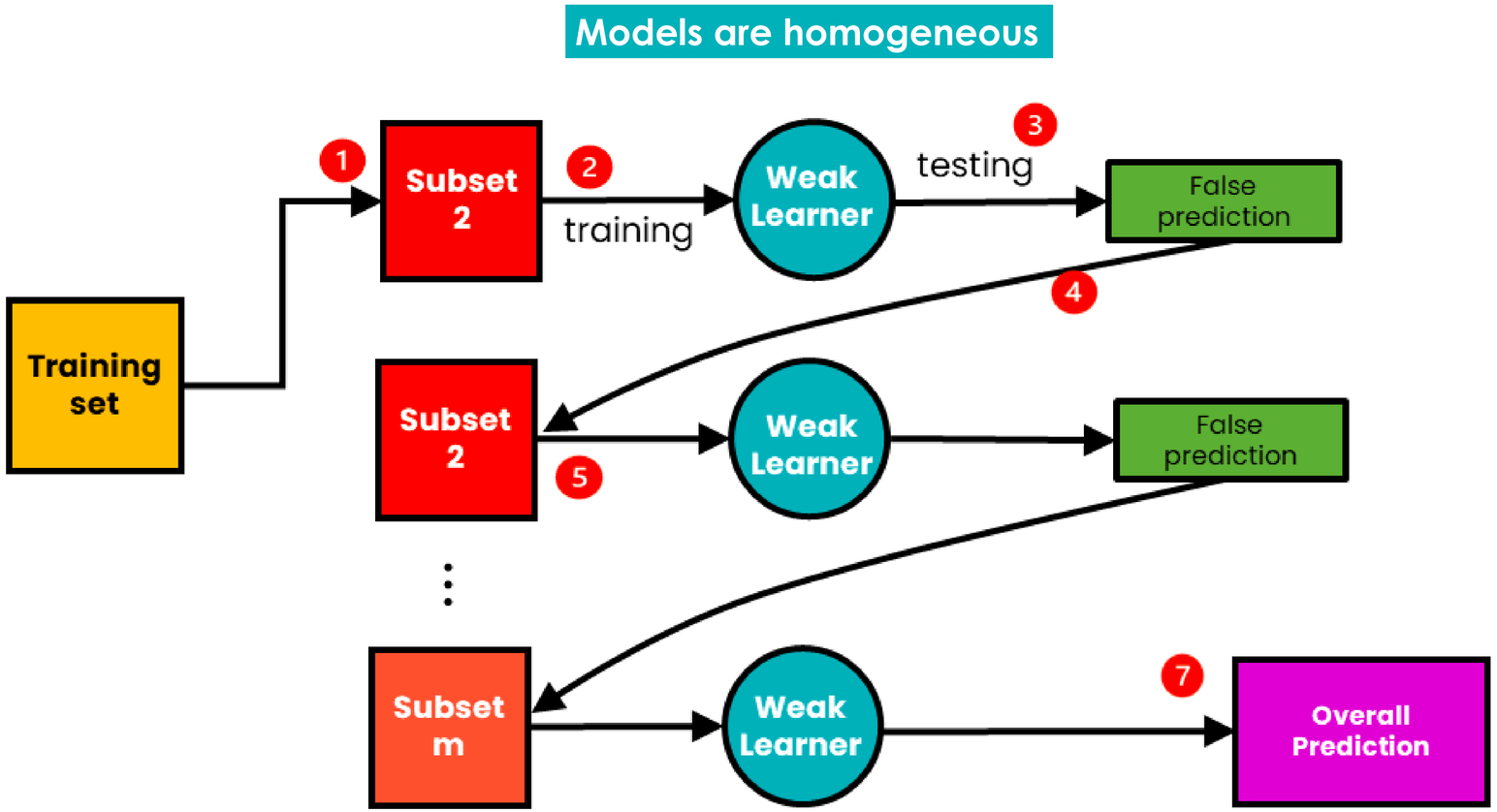


stackfuel

We use boosting for combining weak learners with high bias. Boosting aims to produce a model with a lower bias than that of the individual models

Ensemble Learning Methods

Boosting



First Subset is taken from initial dataset. Weak Learner is trained on the subset.

Incorrectly predicted datapoints are added to the second subset. Models are trained sequentially until final subset is reached.

each subsequent learner improves the errors of previous learners in the sequence

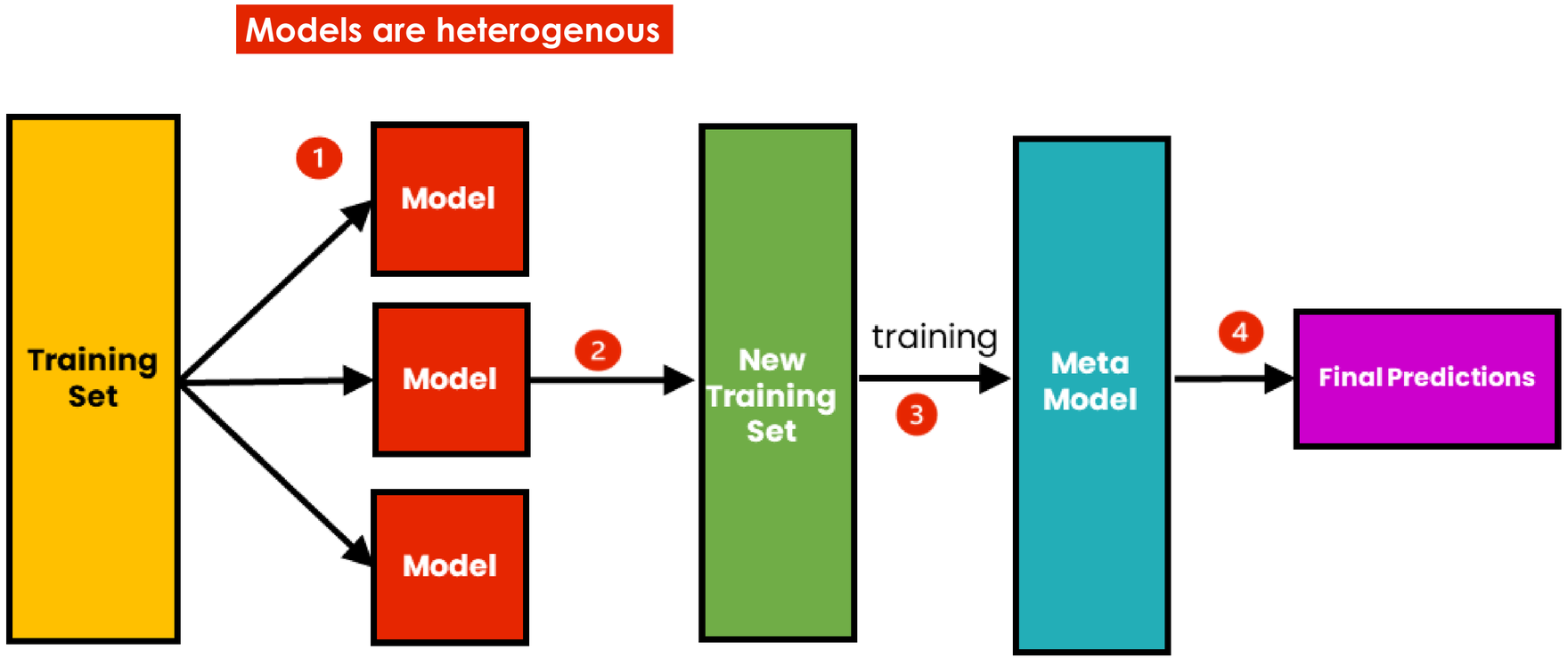
Resulting prediction has been aggregated over the steps (no aggregation necessary anymore)

stackfuel

We use stacking to improve the prediction accuracy of strong learners. Stacking aims to create a single robust model from multiple heterogeneous strong learners.

Ensemble Learning Methods

Stacking



Models are trained on initial dataset and make their predictions

Predictions are then used as features for a metamodel.

Metamodel makes final prediction using weighted averaging.

Strong Learners like bagged and boosted models can be stacked.

stackfuel

Ensemble Learning Methods
Differences

	Bagging	Boosting	Stacking
Purpose	Reduce Variance	Reduce Bias	Improve Accuracy
Base Learner Types	Homogeneous	Homogeneous	Heterogeneous
Base Learner Training	Parallel	Sequential	Meta Model
Aggregation	Max Voting, Averaging	Weighted Averaging	Weighted Averaging

Sources and helpful Links

Images:

- slide 5: https://www.tibco.com/de/reference-center/what-is-a-random-forest
- slide 4,6,7,8:
- https://www.analyticsvidhya.com/blog/2023/01/ensemble-learning-methods-bagging-boosting-and-stacking/