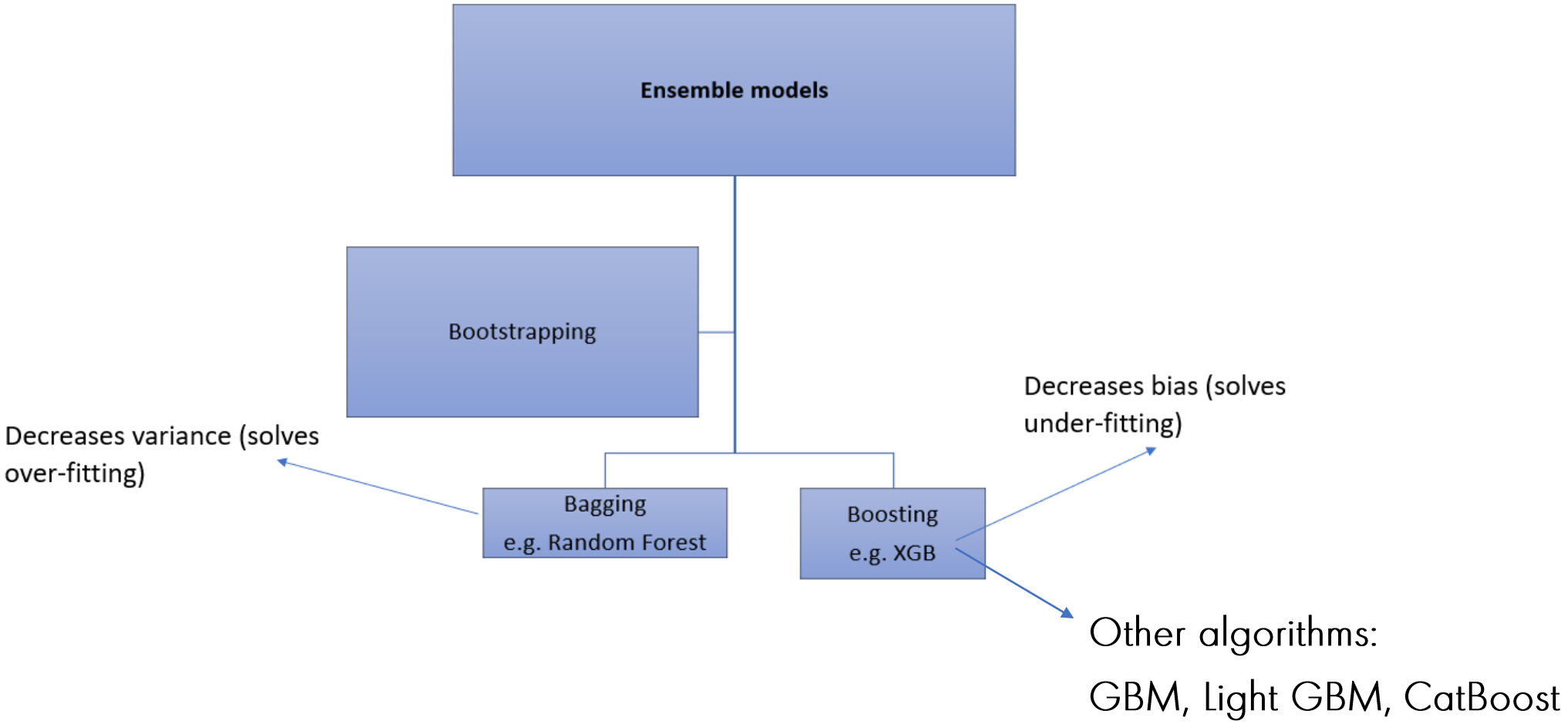
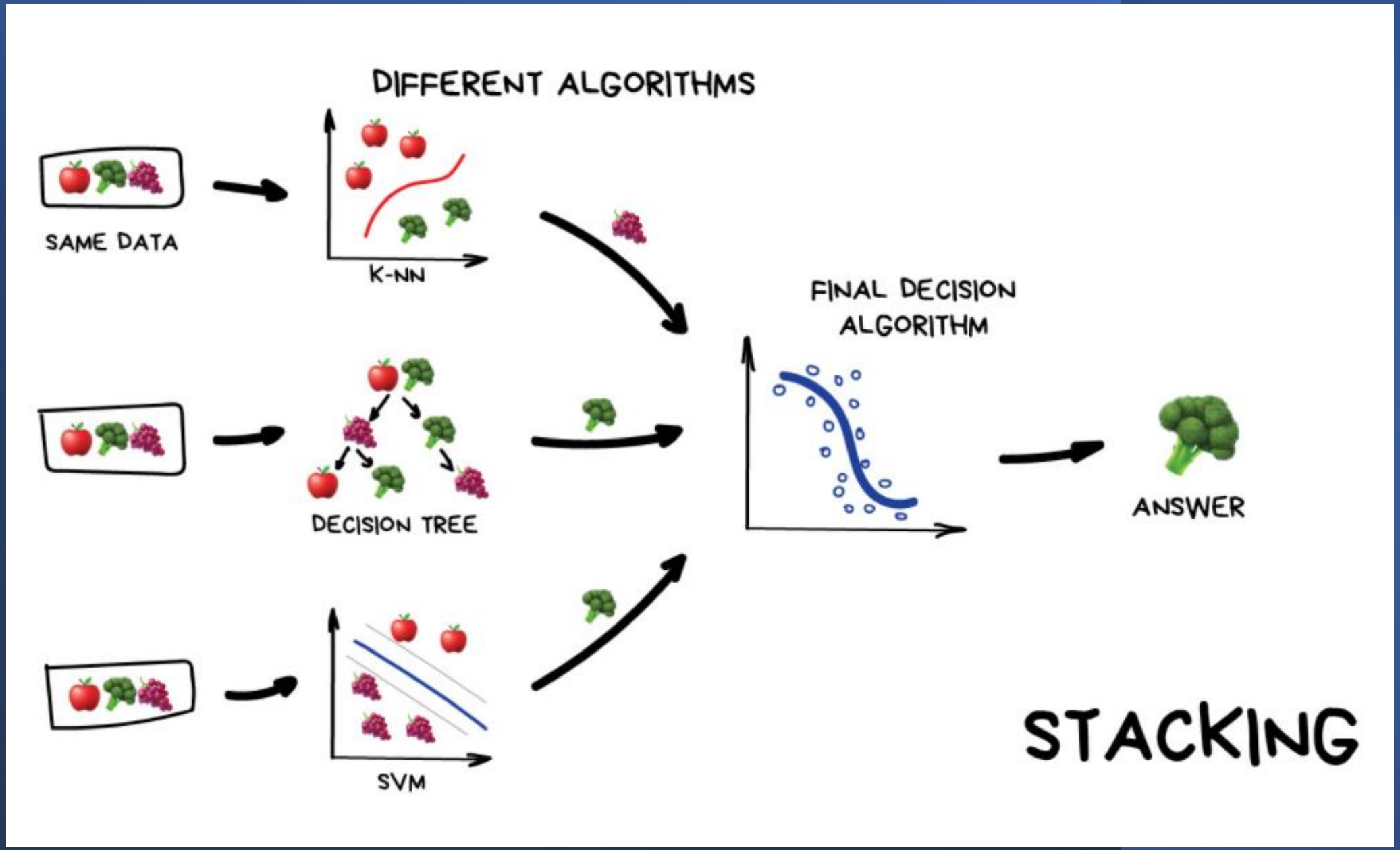
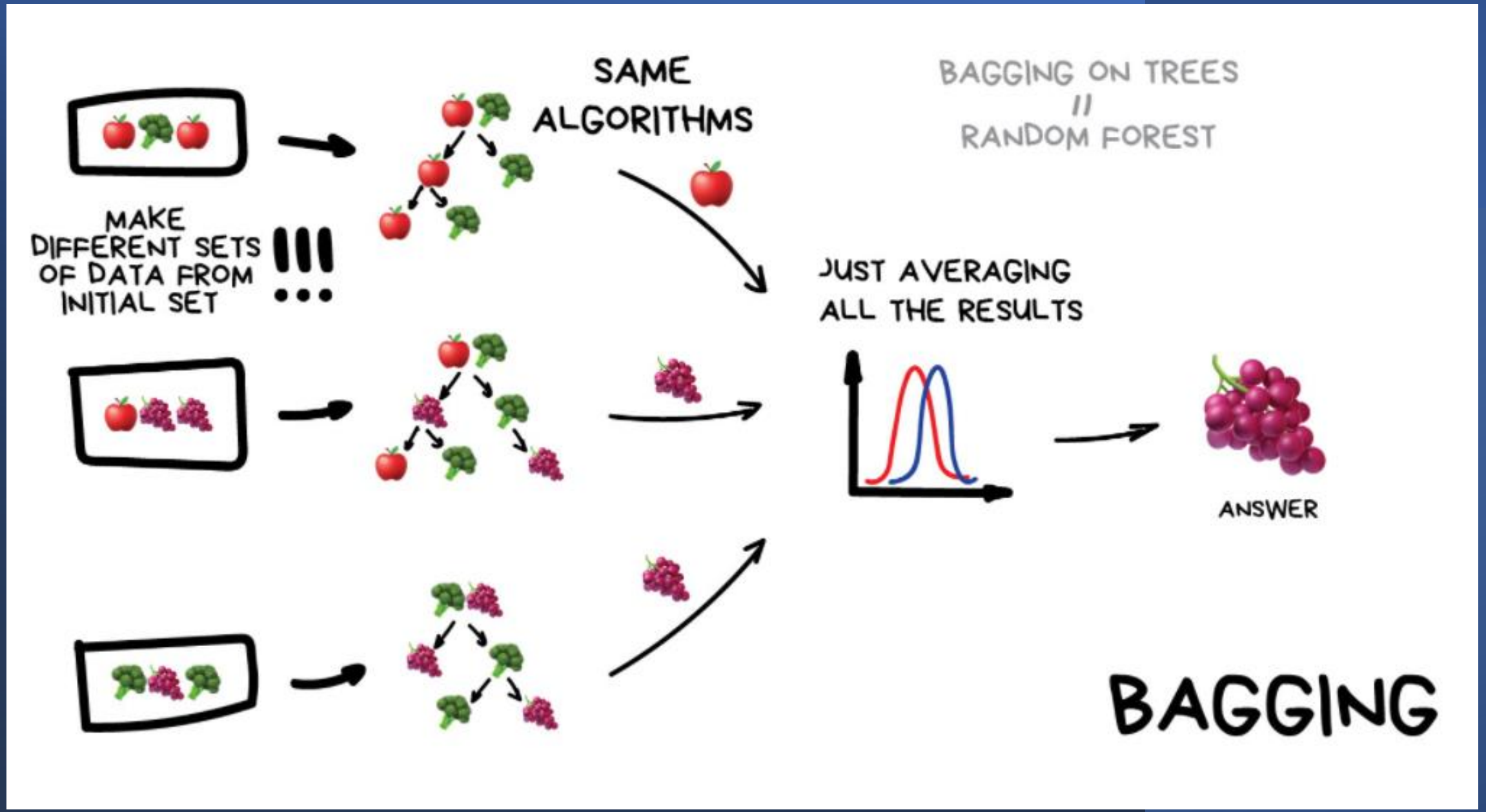
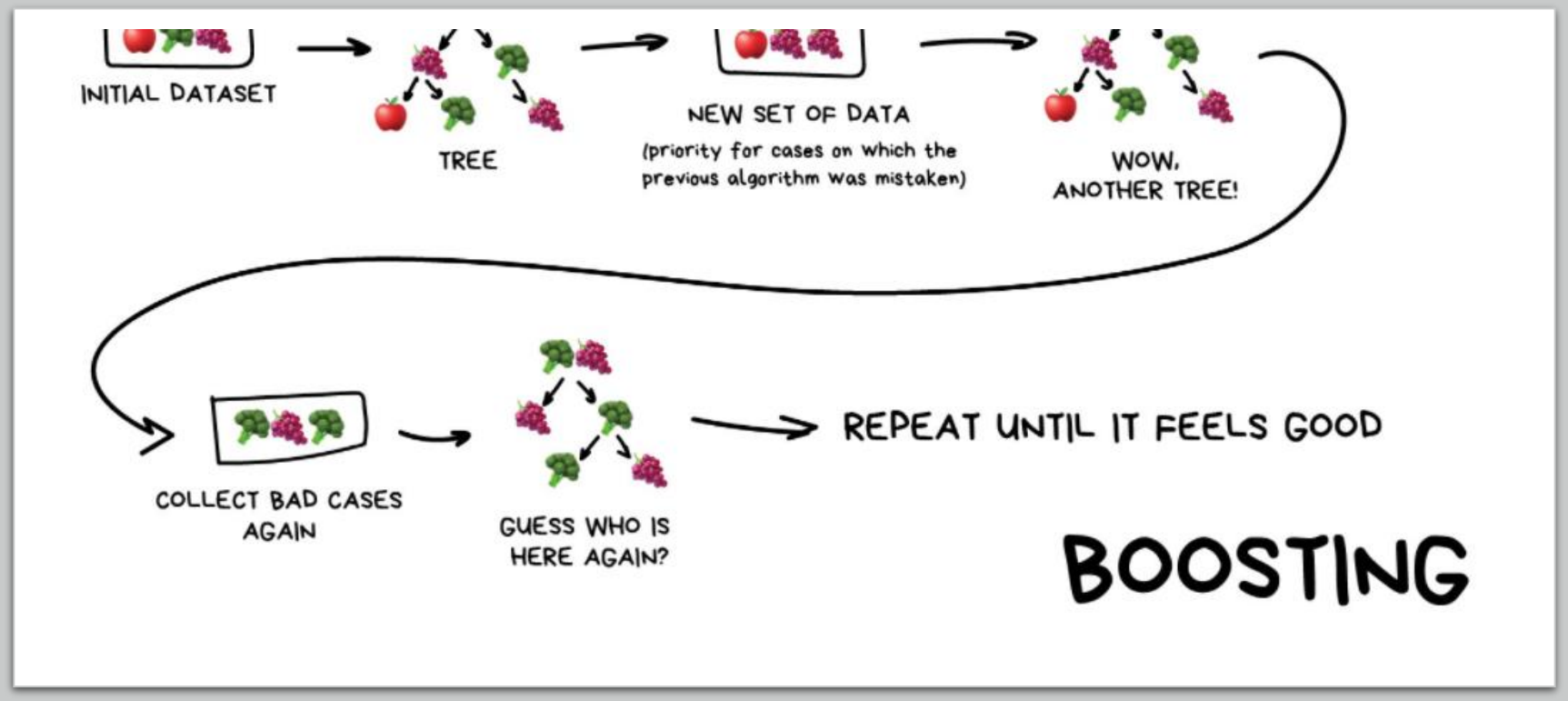
Ensembles


Ensemble models
Bootstrapping
Decreases variance (solves over-fitting)
Bagging
e.g. Random Forest
Boosting
e.g. XGB
Decreases bias (solves under-fitting)
Other algorithms:
GBM, Light GBM, CatBoost


DIFFERENT ALGORITHMS
SAME DATA
K-NN
DECISION TREE
SVM
FINAL DECISION ALGORITHM
ANSWER
STACKING


SAME ALGORITHMS
BAGGING ON TREES
=
RANDOM FOREST
MAKE DIFFERENT SETS OF DATA FROM INITIAL SET
!!!
JUST AVERAGING ALL THE RESULTS
ANSWER
BAGGING



- For mixed datatypes, tree-based models are a good choice. Random Forest is one choice of baseline prediction model, they do not need extensive hyperparameter tuning.