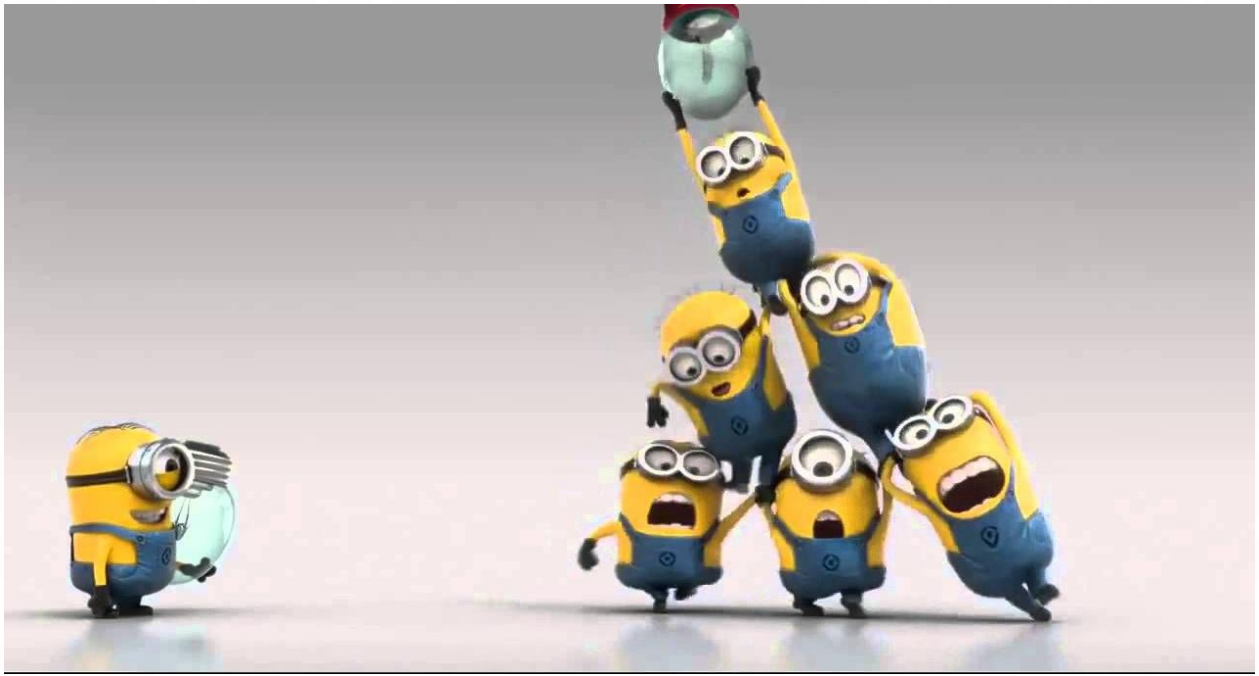
ENSEMBLE LEARNING

JENS BAETENS



https://www.youtube.com/watch?v=Un9zObFjBH0

KENMERKEN

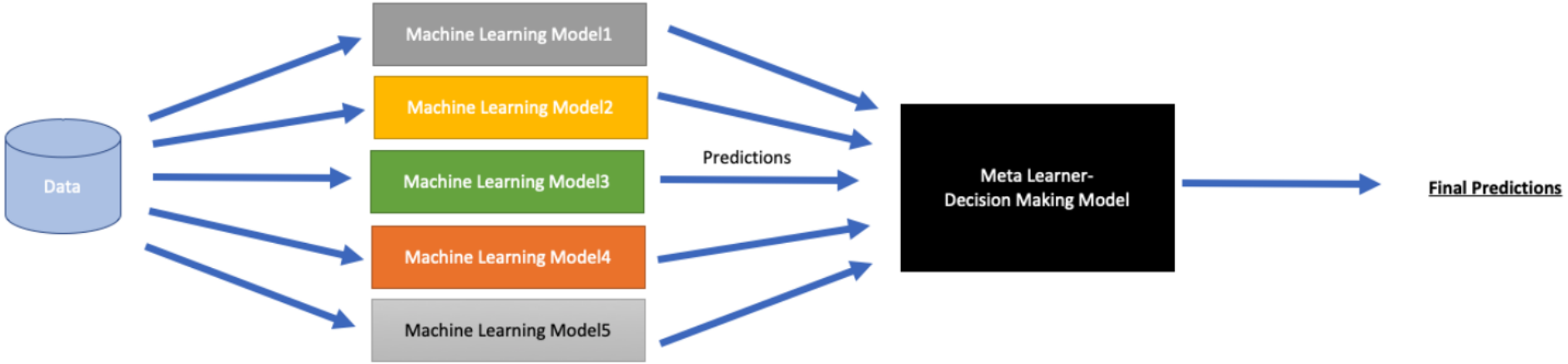
Meerdere modellen die samen de correcte output proberen te voorspelen

Gaat zowel voor classificatie als regressie

Meerdere technieken kunnen gecombineerd worden

=> Doel is om een betere voorspelling te bekomen dan de afzonderlijke modellen

STACKING



Decision making model: Majority voting of een nieuwe classifier

Kan op basis van de predicties maar ook op basis van de kansen van elke classifier

https://towardsdatascience.com/ensemble-models-for-classification-d443ebed7efe

BAGGING – BOOTSTRAP AGGREGATING

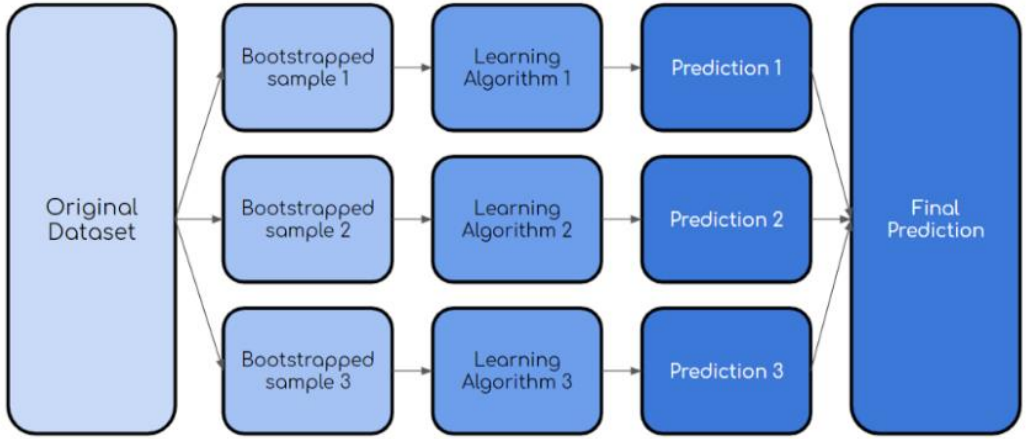
Meerdere modellen van hetzelfde algoritme

Train elk model met deel van de data
(Vaak rond 60%)

Samplen gebeurd met teruglegging

Standaard majority voting

Verminderd overfitting



https://towardsdatascience.com/ensemble-learning-bagging-and-boosting-explained-in-3-minutes-2e6d2240ae21

BOOSTING

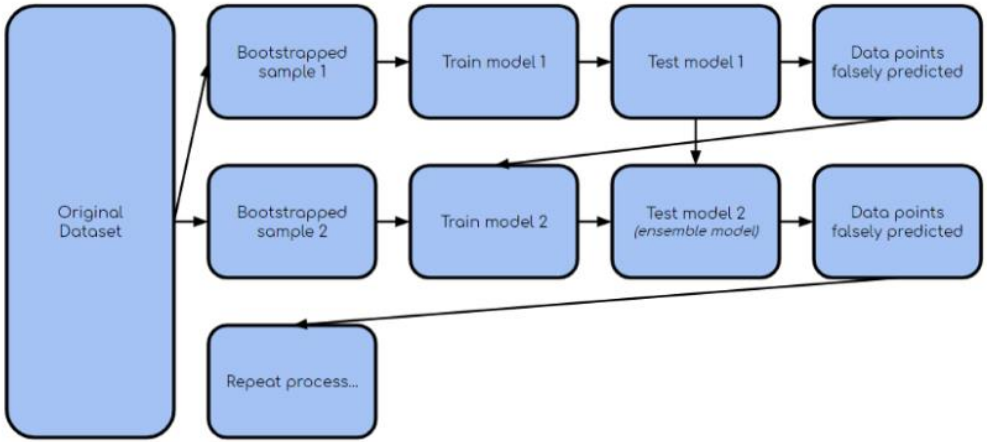
Gelijkaardig aan bagging

Maar training moet sequentieel gebeuren

Misgeclassificeerde samples grotere kans om gebruikt te worden om het volgende model te trainen

Meest gekende algoritme: AdaBoosting

VB: Face detection in camera's



https://towardsdatascience.com/ensemble-learning-bagging-and-boosting-explained-in-3-minutes-2e6d2240ae21

XGBOOST

https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/

Heel snel om te trainen

Haalt goede scores

https://www.youtube.com/watch?v=VgDe0gwesrw

Features of XGBoost

- Regularized boosting (prevents overfitting)
- Can handle missing values automatically
- Parallel processing
- Can cross-validate at each iteration
  - Enables early stopping, finding optimal number of iterations
- Incremental training
- Can plug in your own optimization objectives
- Tree pruning
  - Generally results in deeper, but optimized, trees

https://www.youtube.com/watch?v=OQKQHNCVf5k