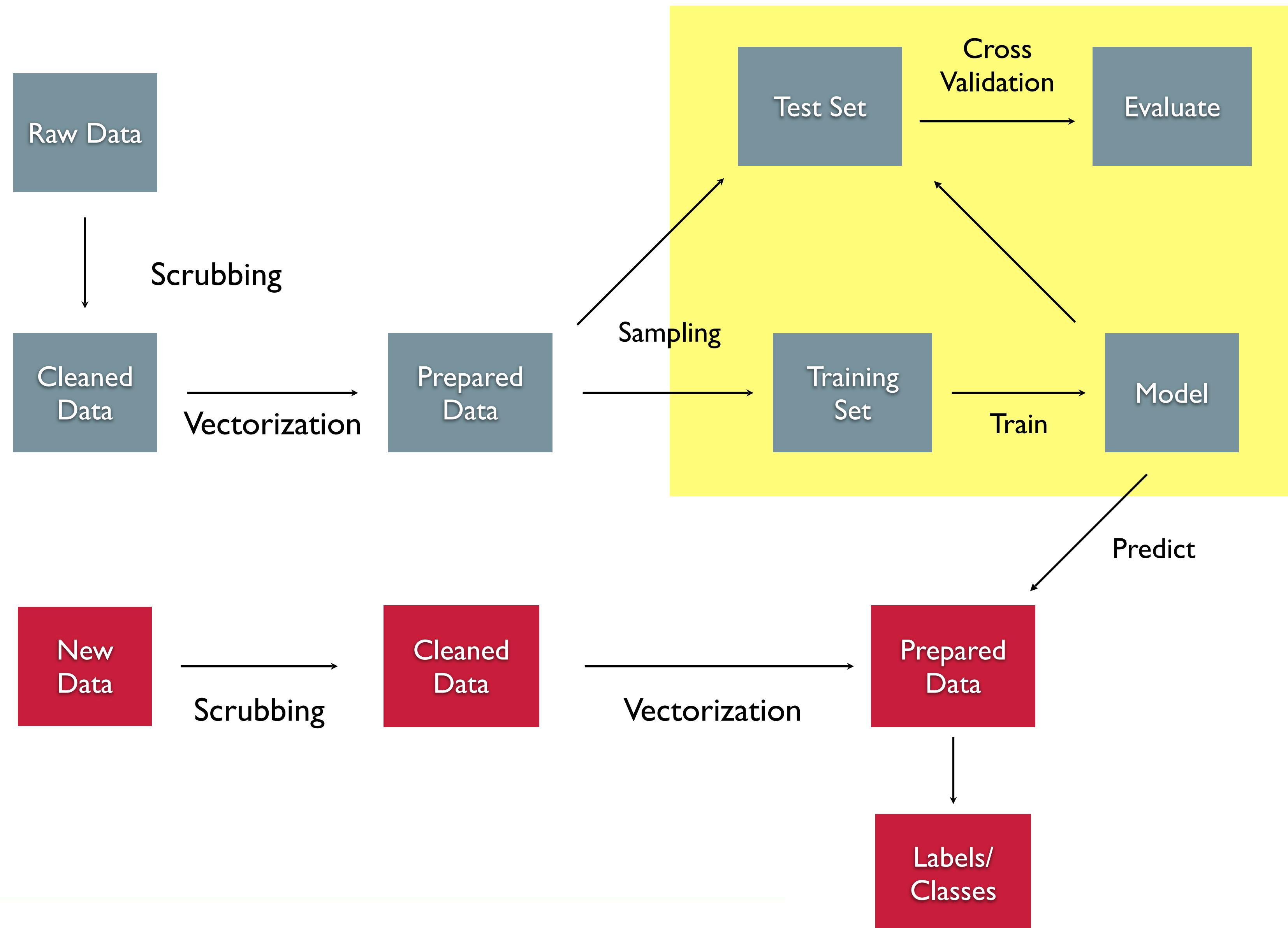


Overview

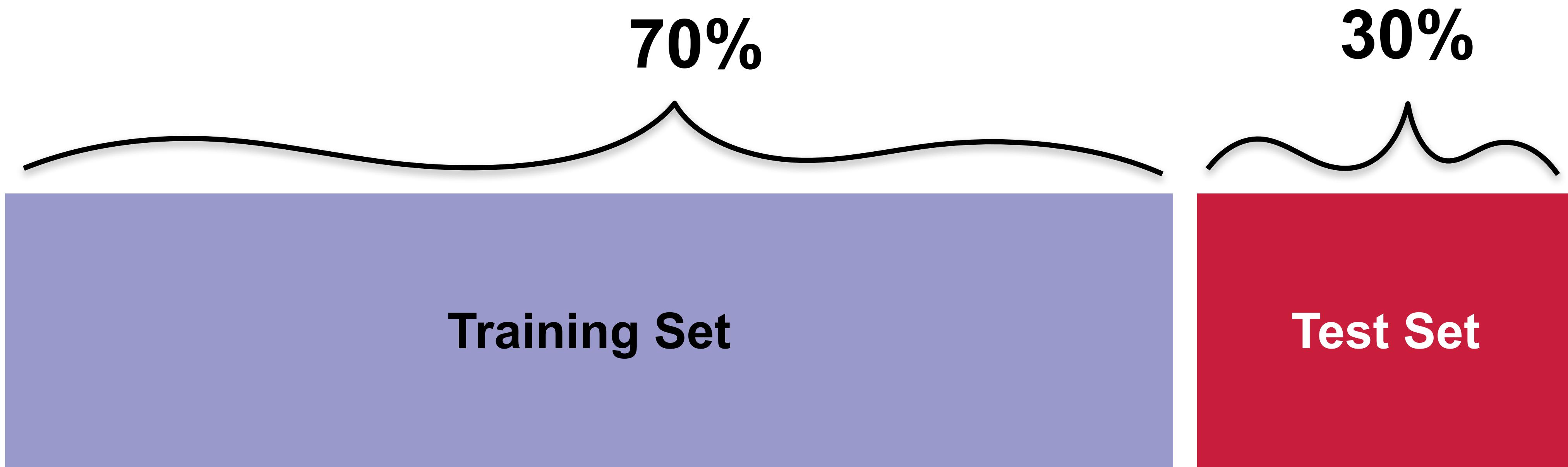


Evaluation

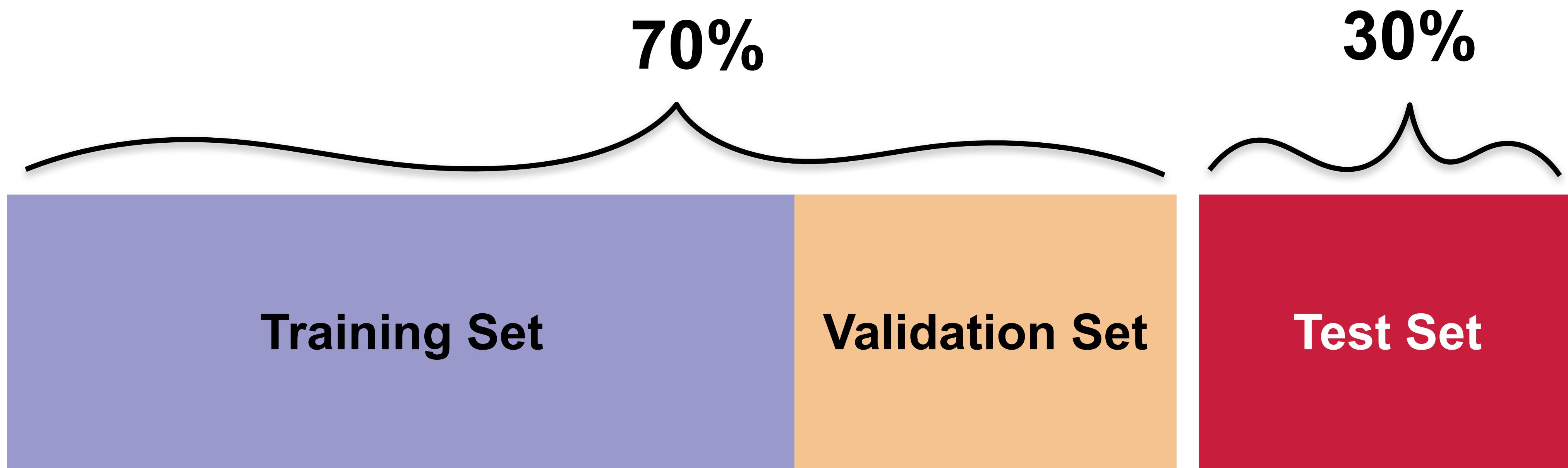
Split labeled data into a **training set** and a **test set**

Whatever you do, **DO NOT** cross the streams

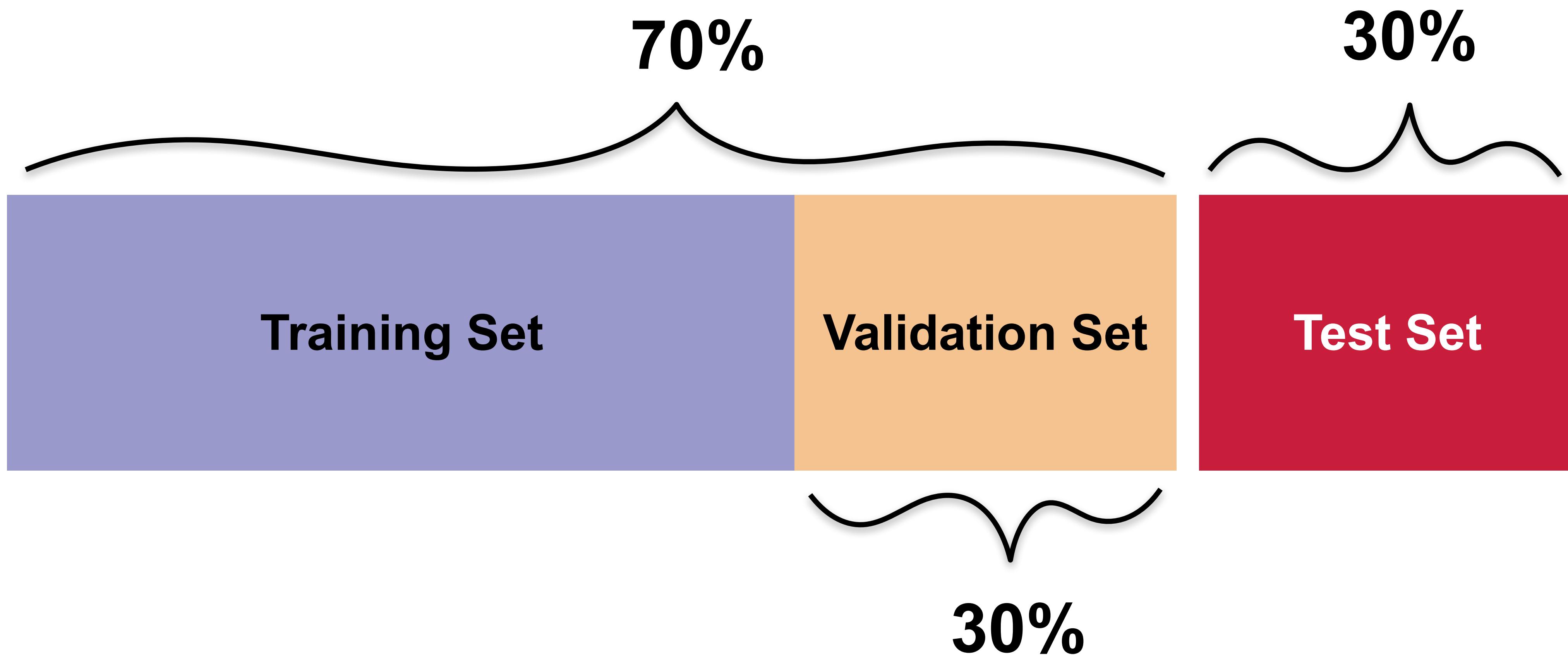
Holdout Cross Validation



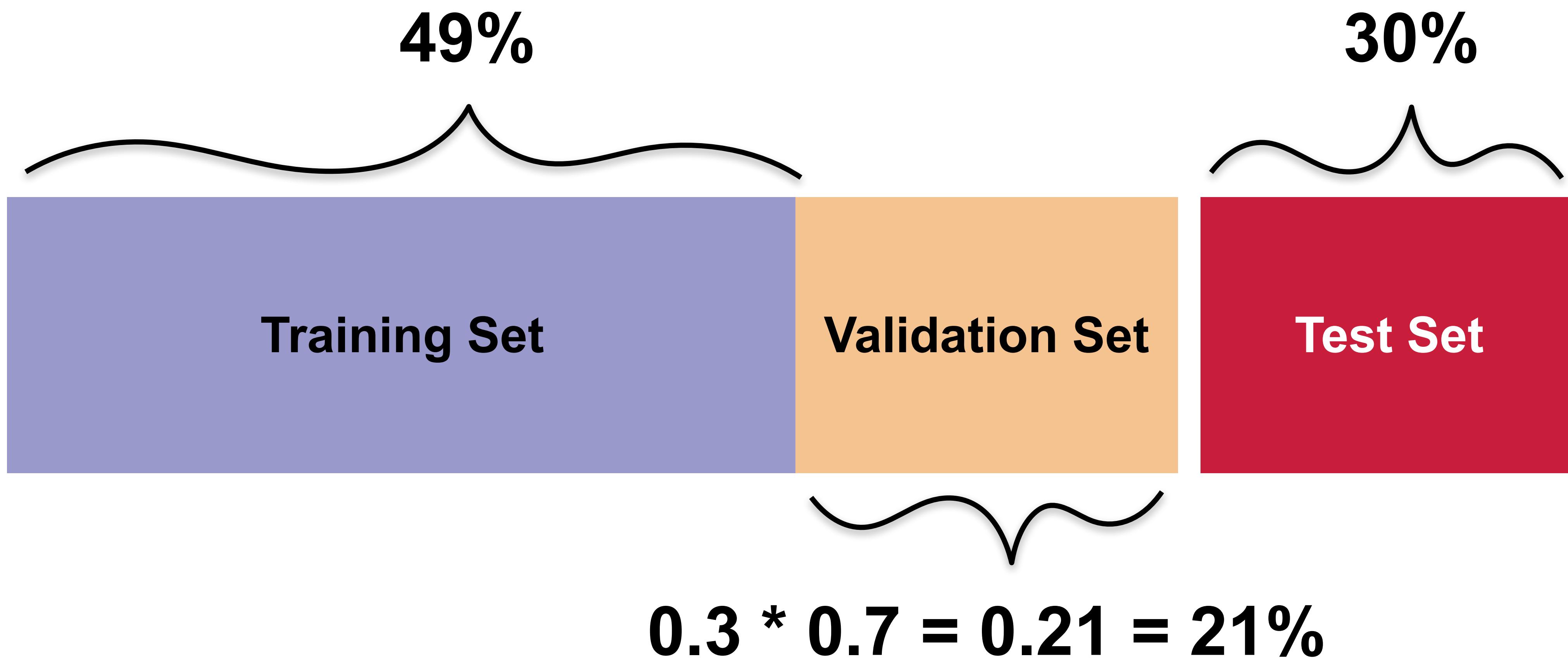
Holdout Cross Validation



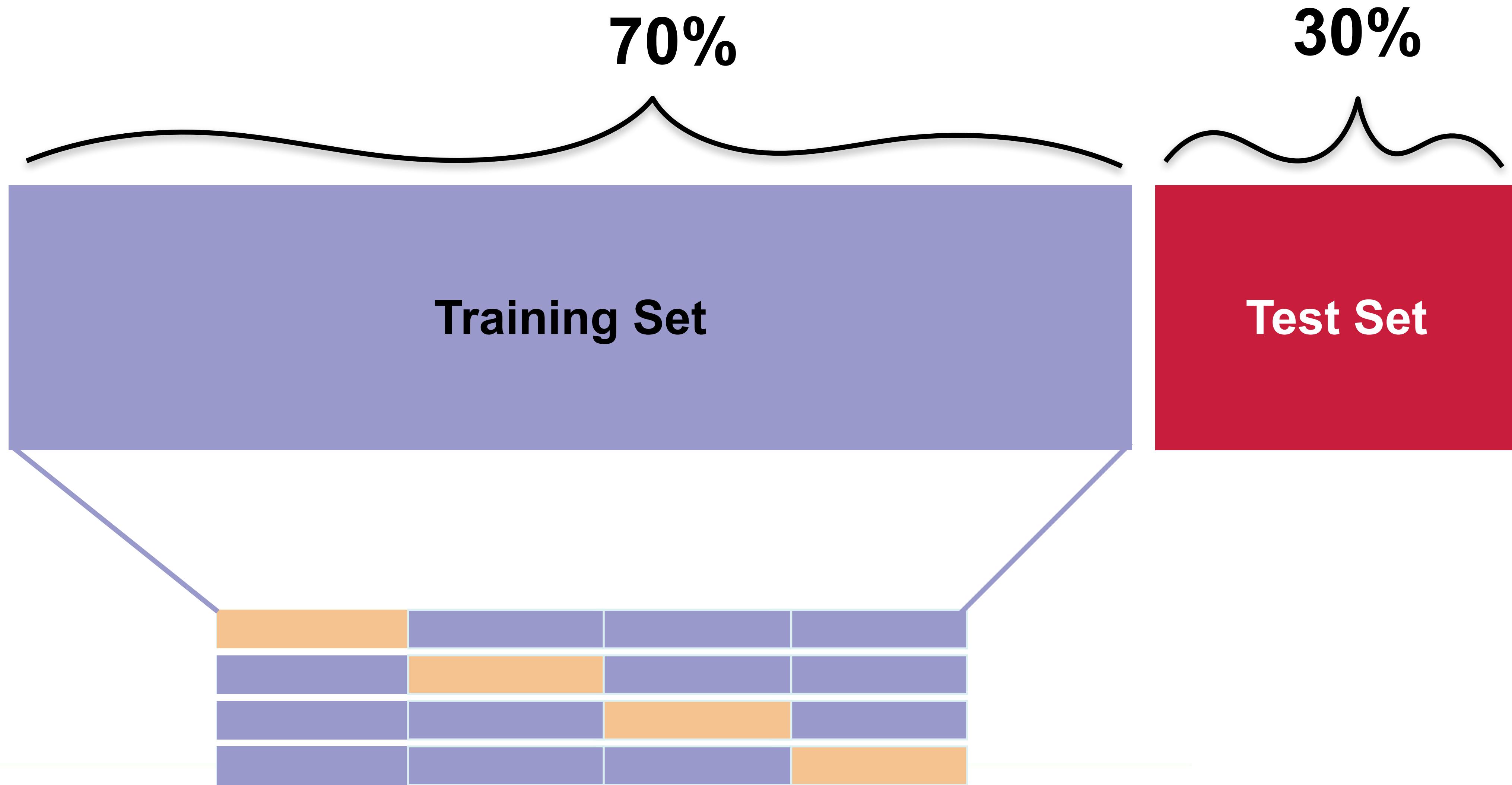
Holdout Cross Validation



Holdout Cross Validation



K-fold Cross Validation



Evaluation Metrics

- Accuracy
- Precision/Recall
- F1
- AUC (area under the curve)



Bias vs. Variance

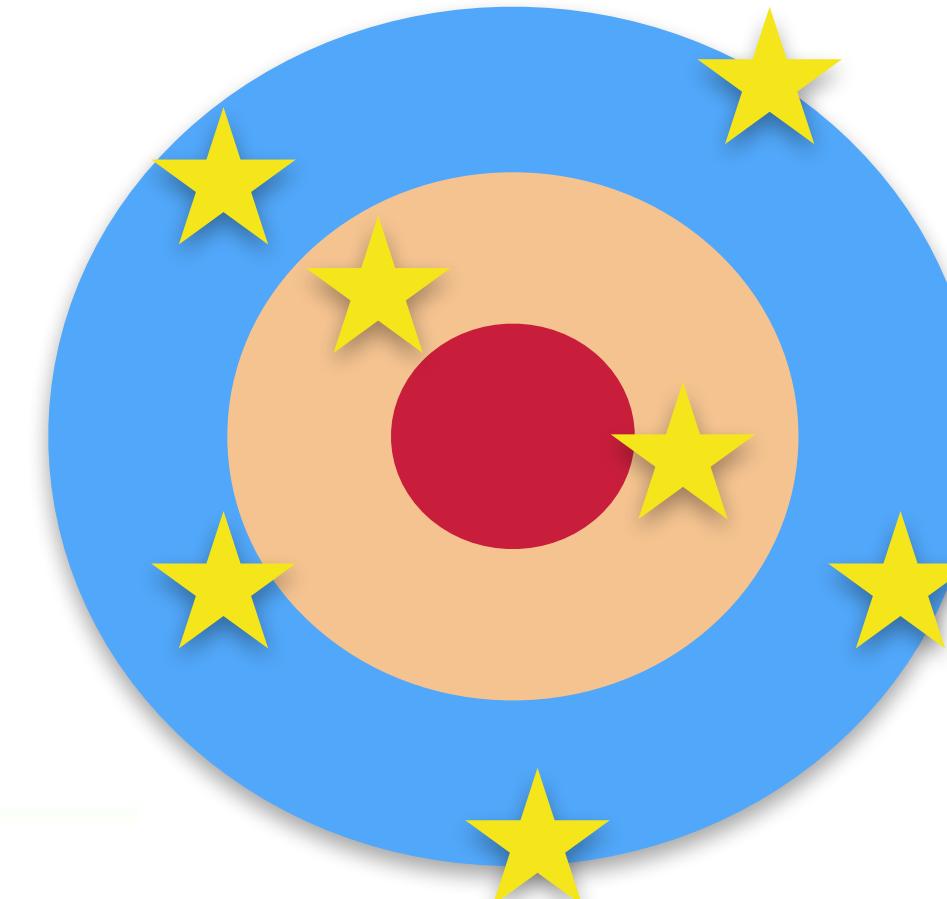
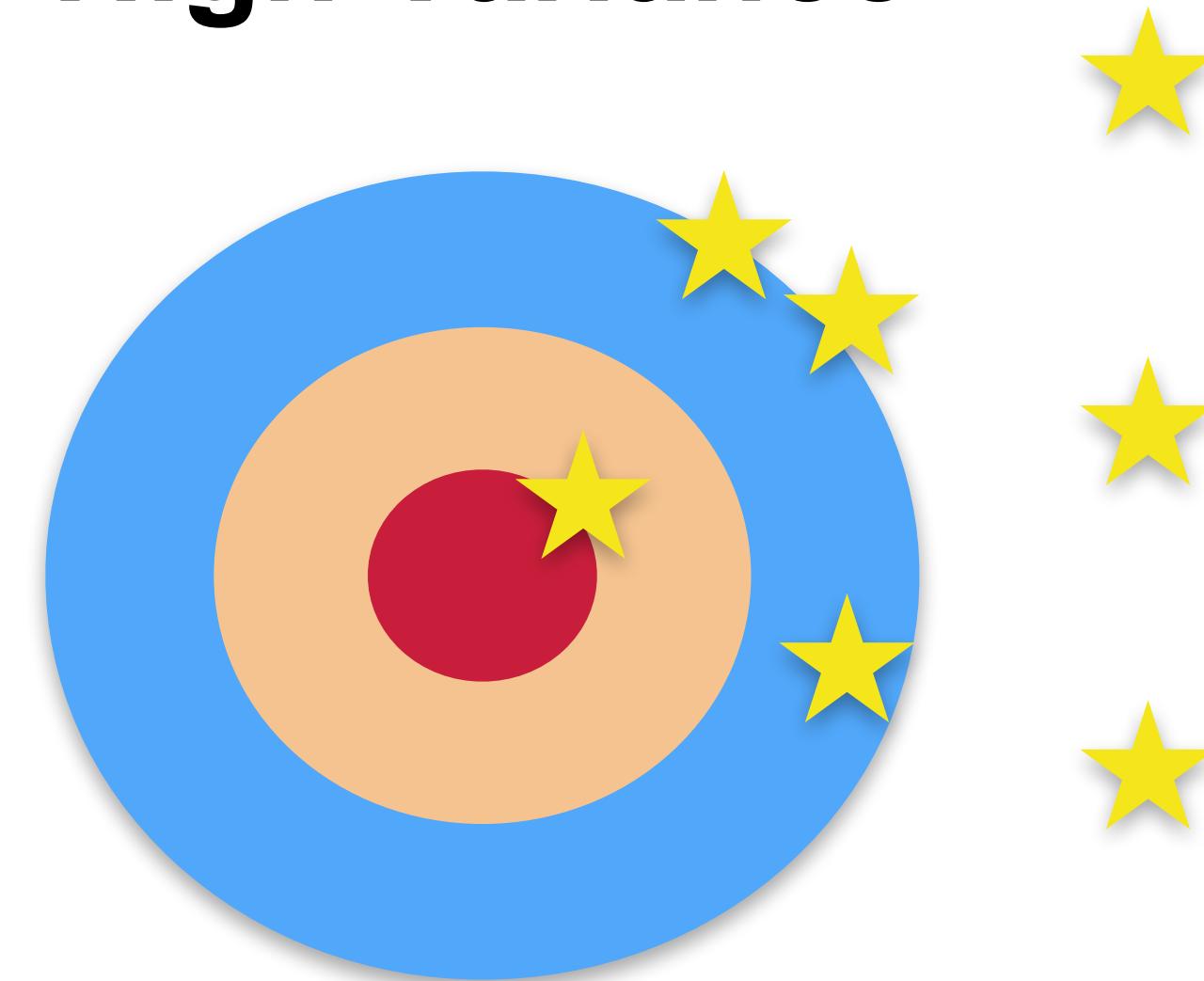
High Bias

Low Variance



Low Bias

High Variance



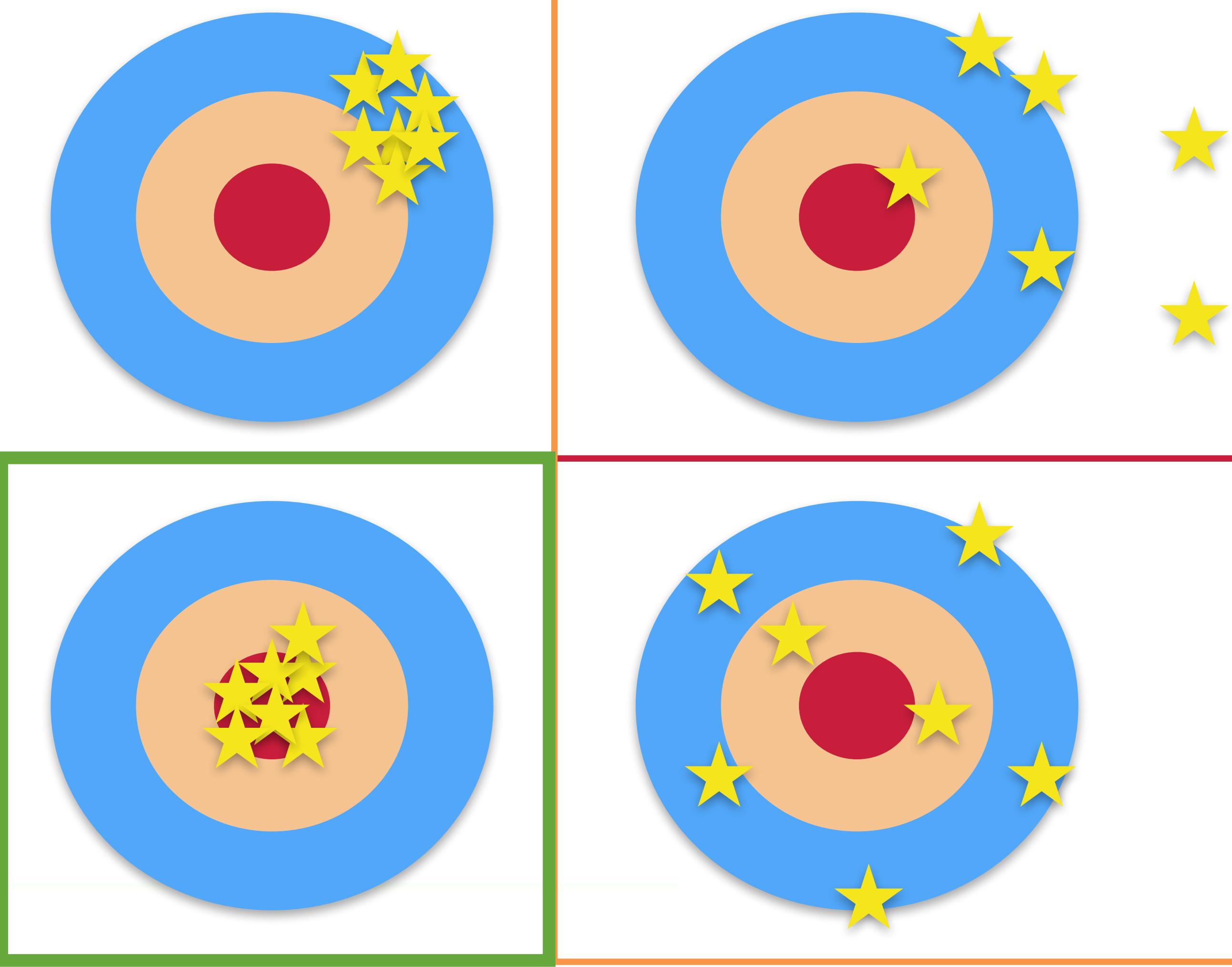
Pearson

Bias vs. Variance

High complexity

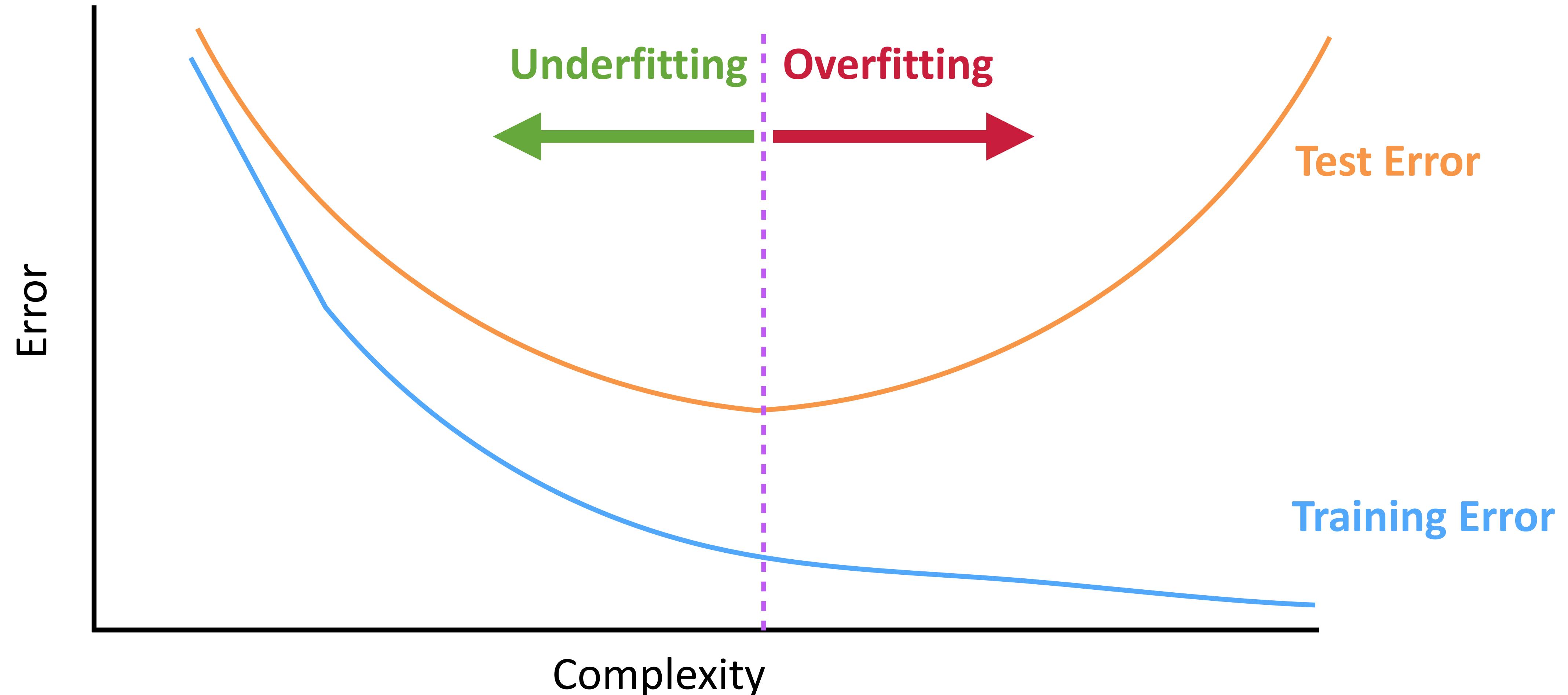
Low complexity

Just right ➡

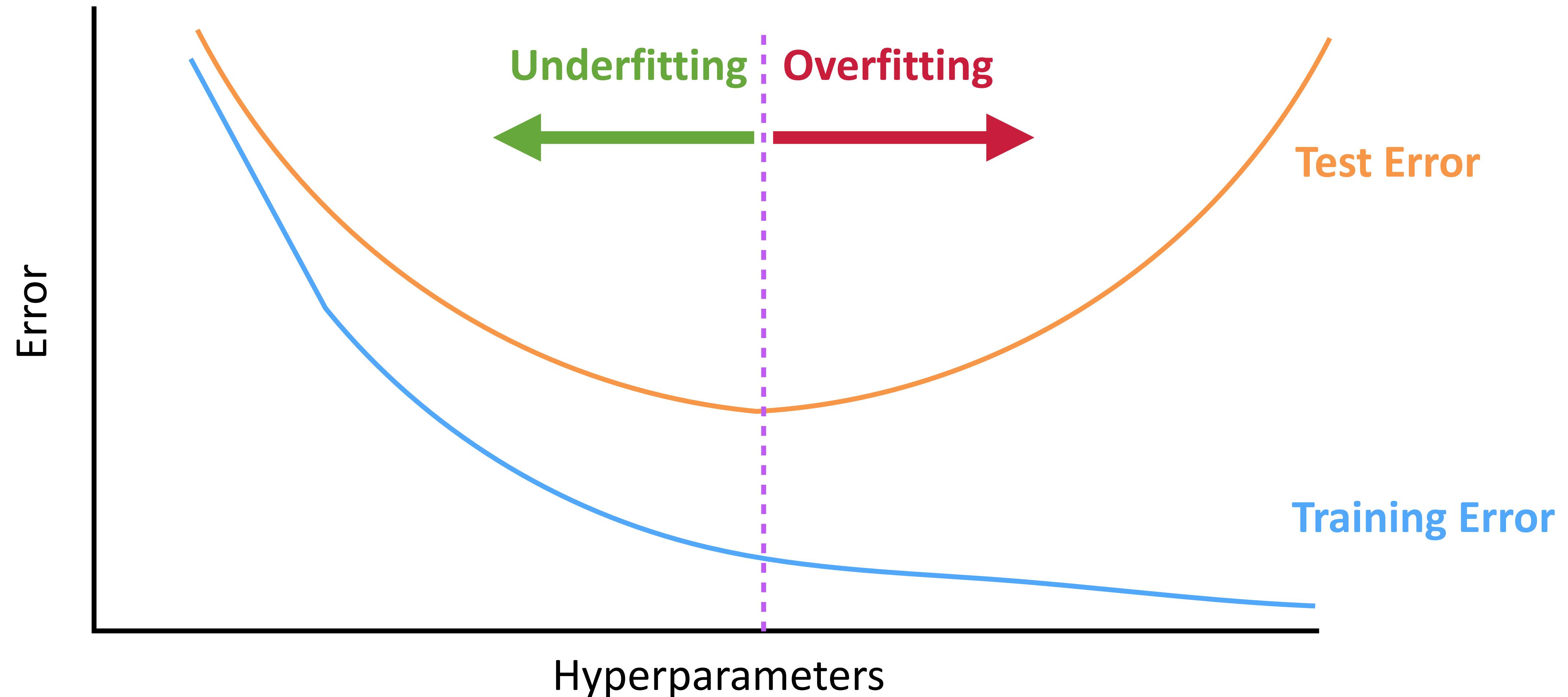


Pearson

Bias-variance trade-off



Bias-variance trade-off



Grid Search

- Exhaustive **brute force** search
- Find optimal **hyperparameters** or models
- Computationally costly
- But **embarrassingly parallel!**



Can we do better?

- Better search through hyperparameter space
- Better resource allocation and utilization



Can we do better?

- Bayesian Optimization
- Early Stopping

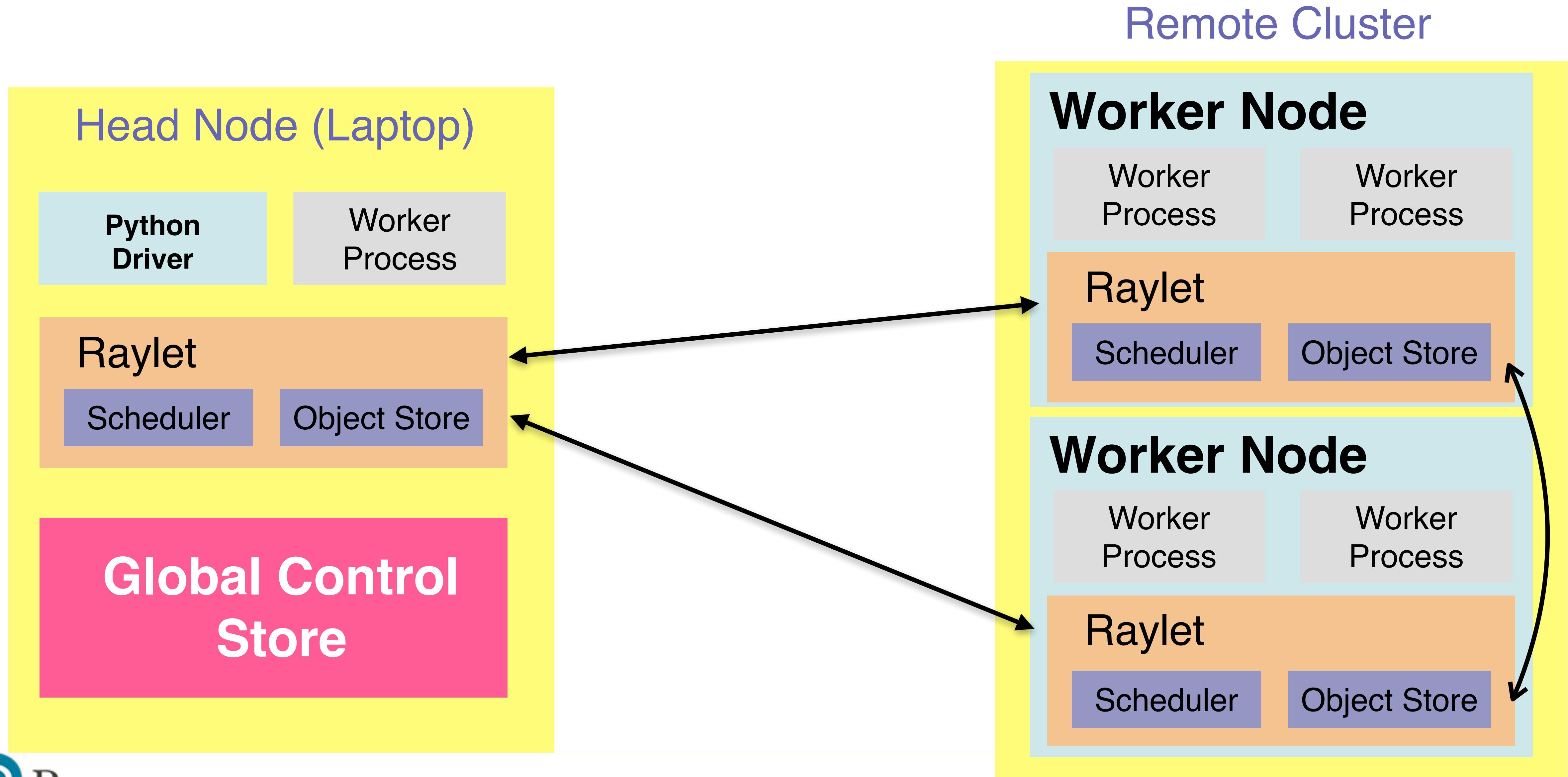


Can we do better?

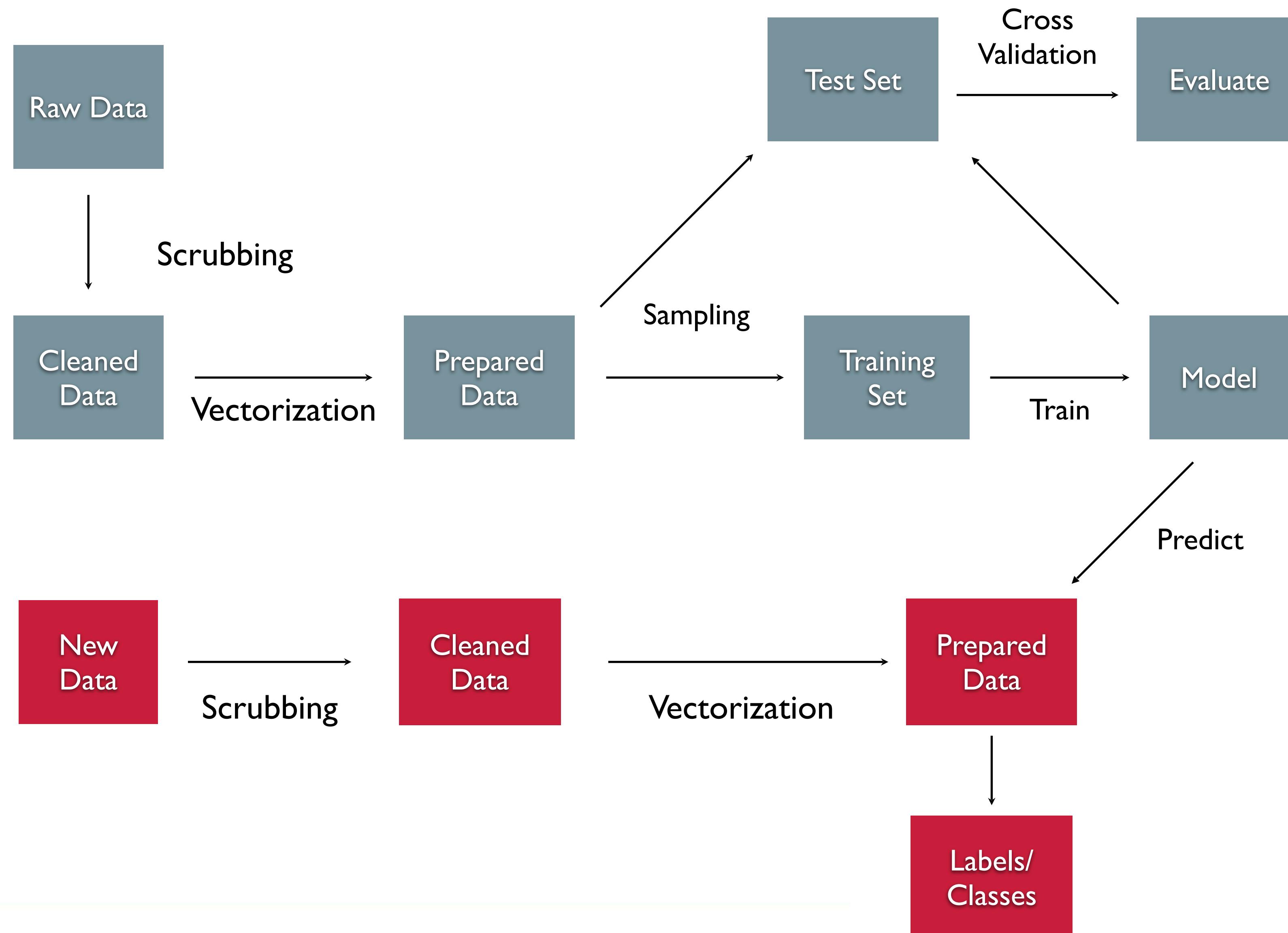
- Bayesian Optimization
- Early Stopping



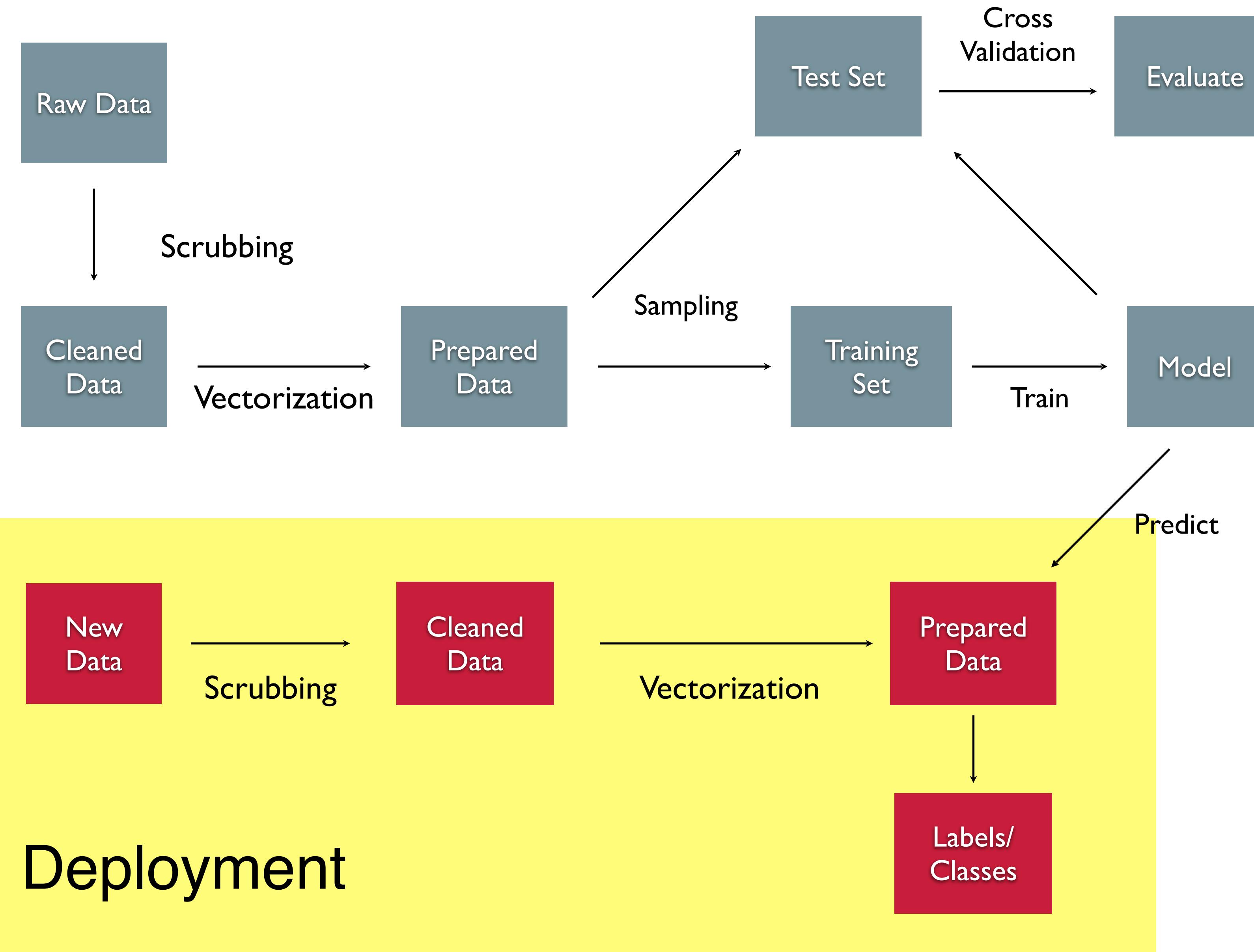
Ray Core Internals



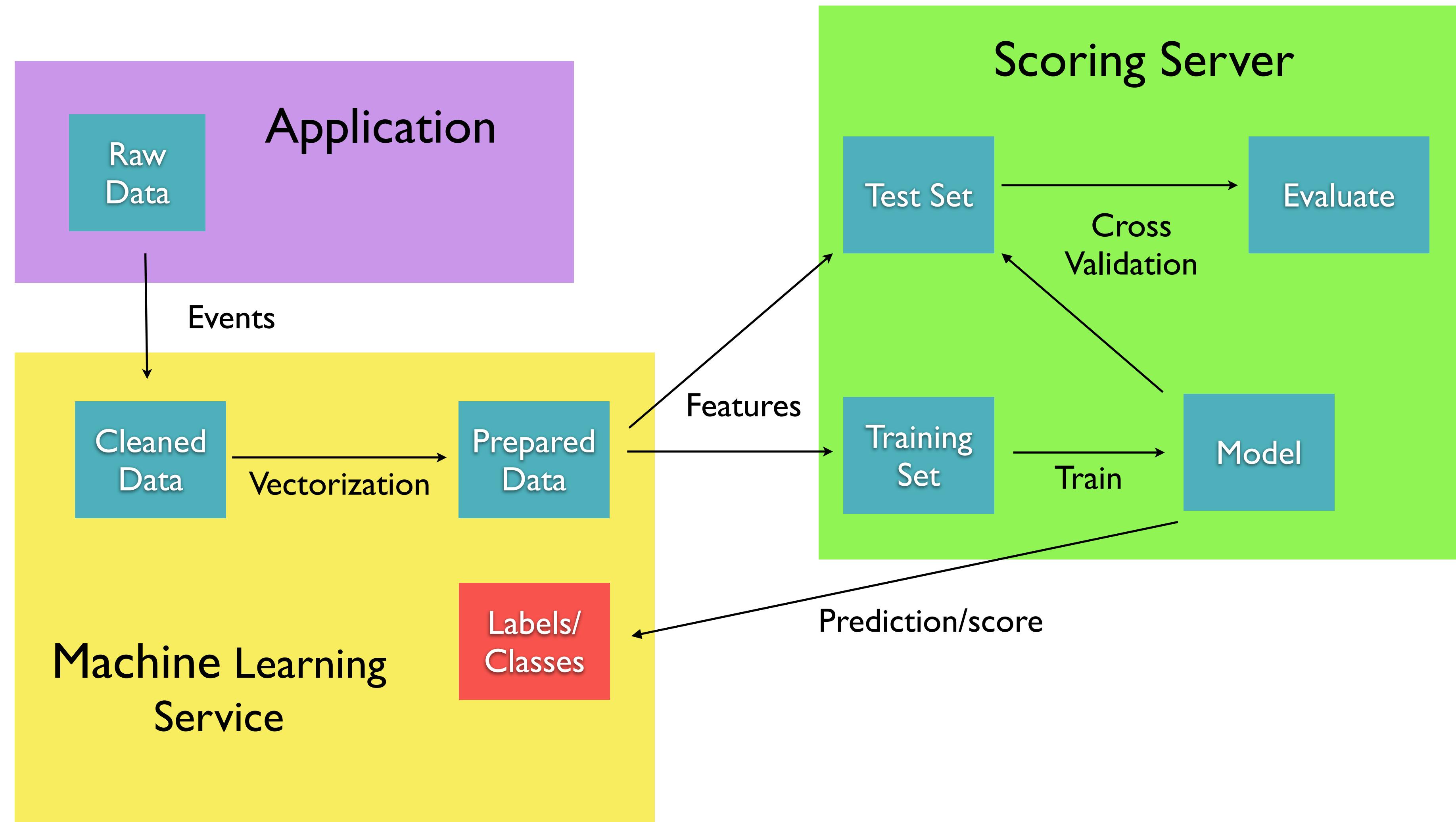
Overview



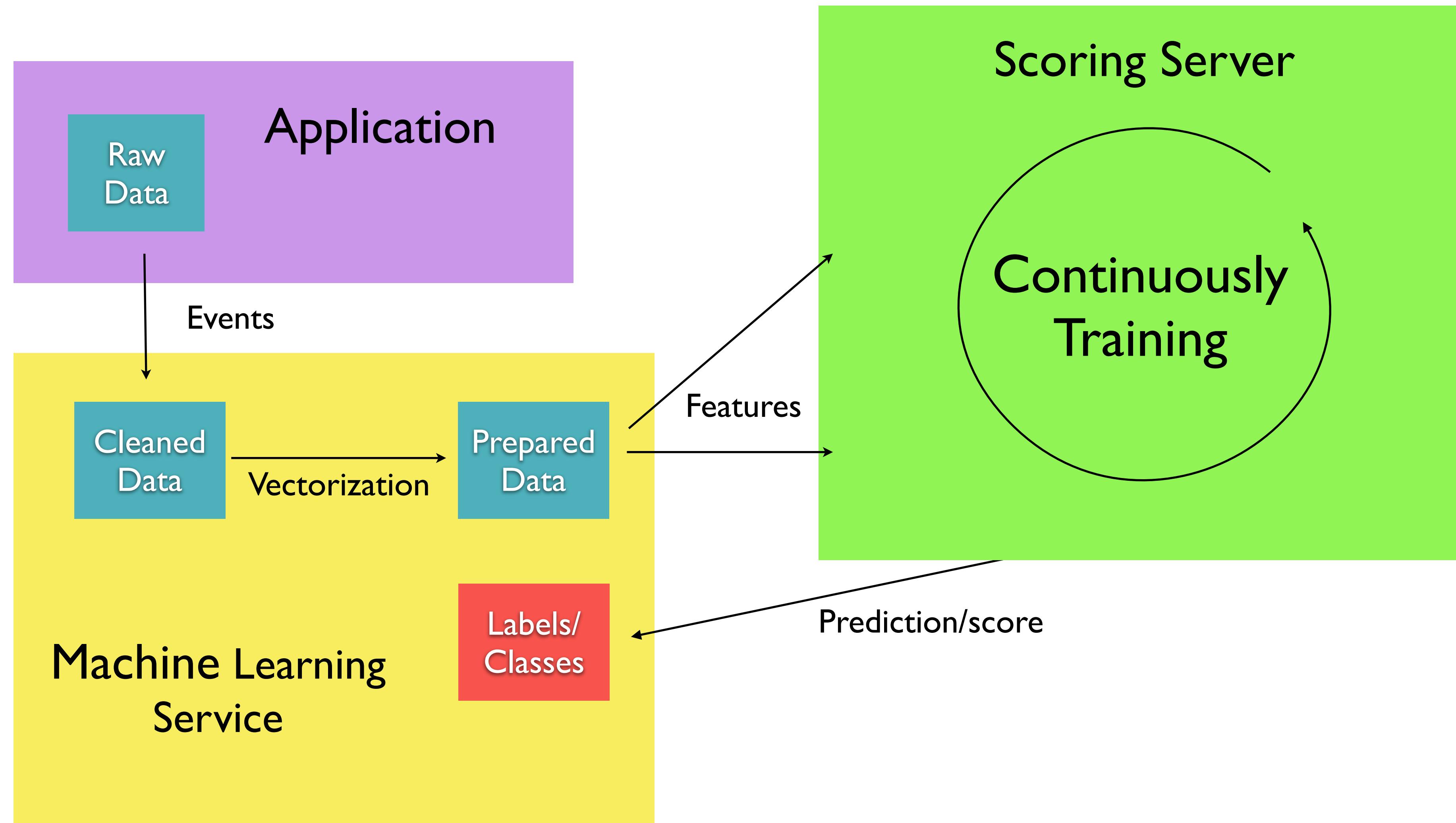
Overview



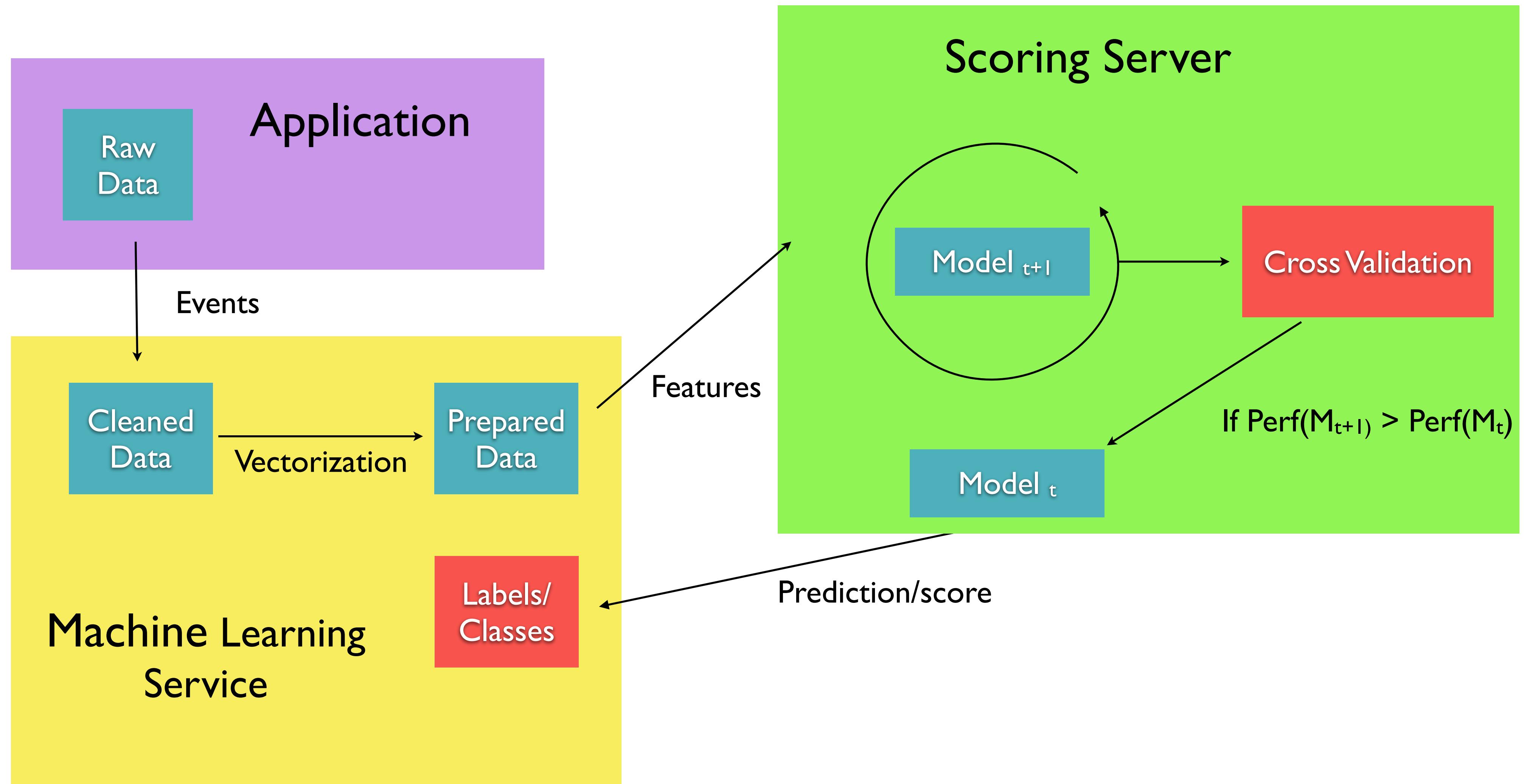
Deploying Models

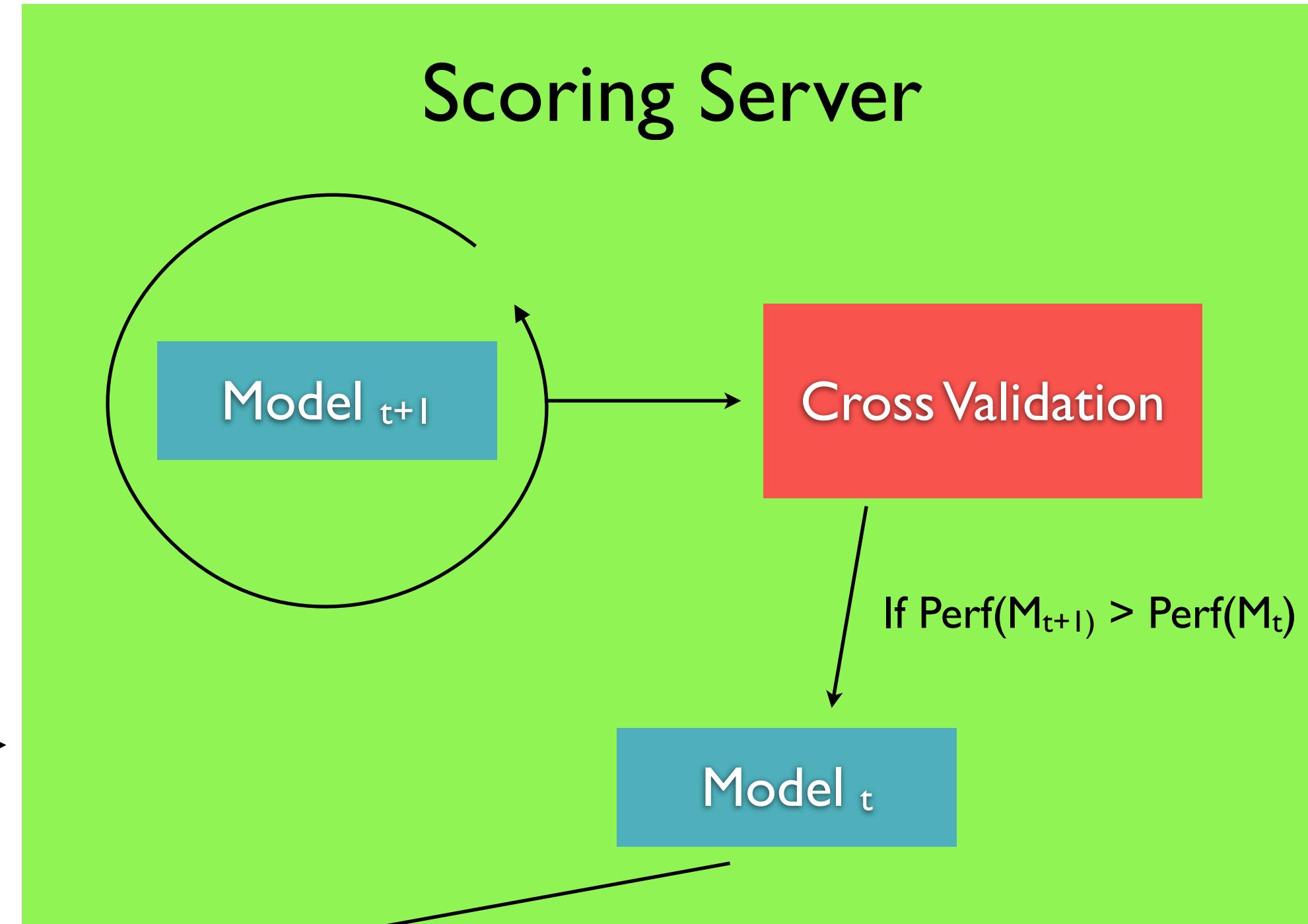
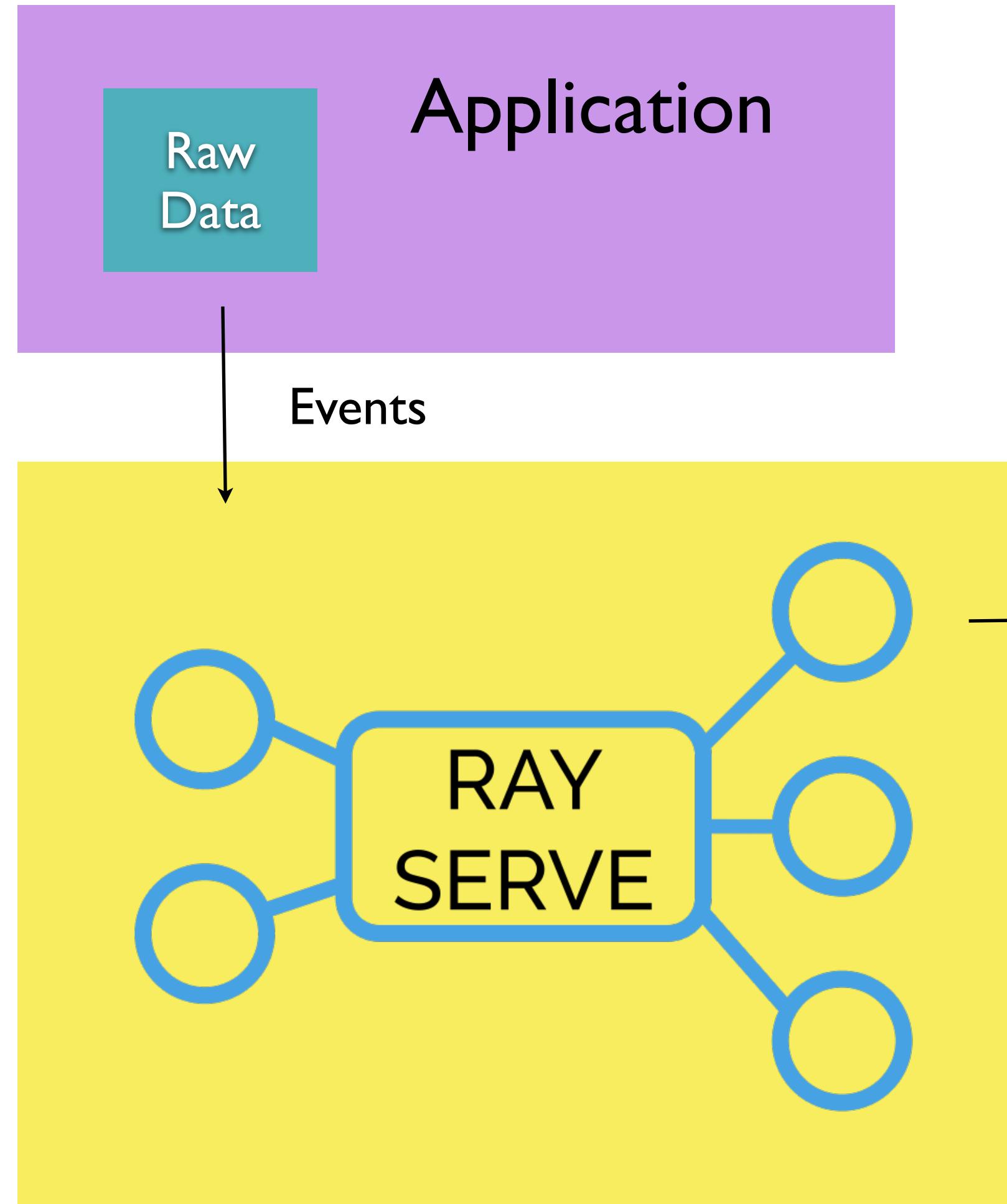


Deploying Models

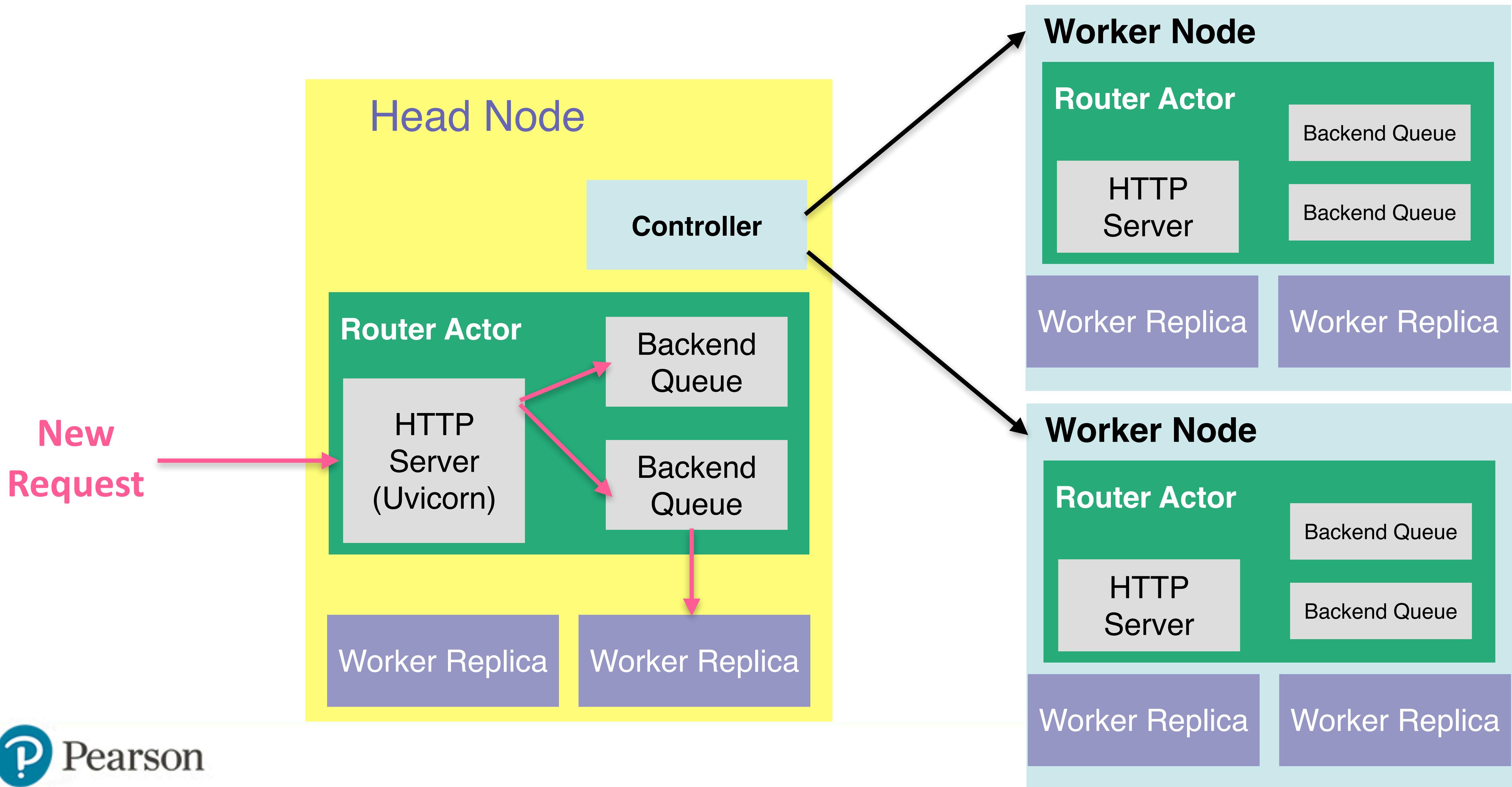


Batch Offline Evaluation





Ray Serve Internals

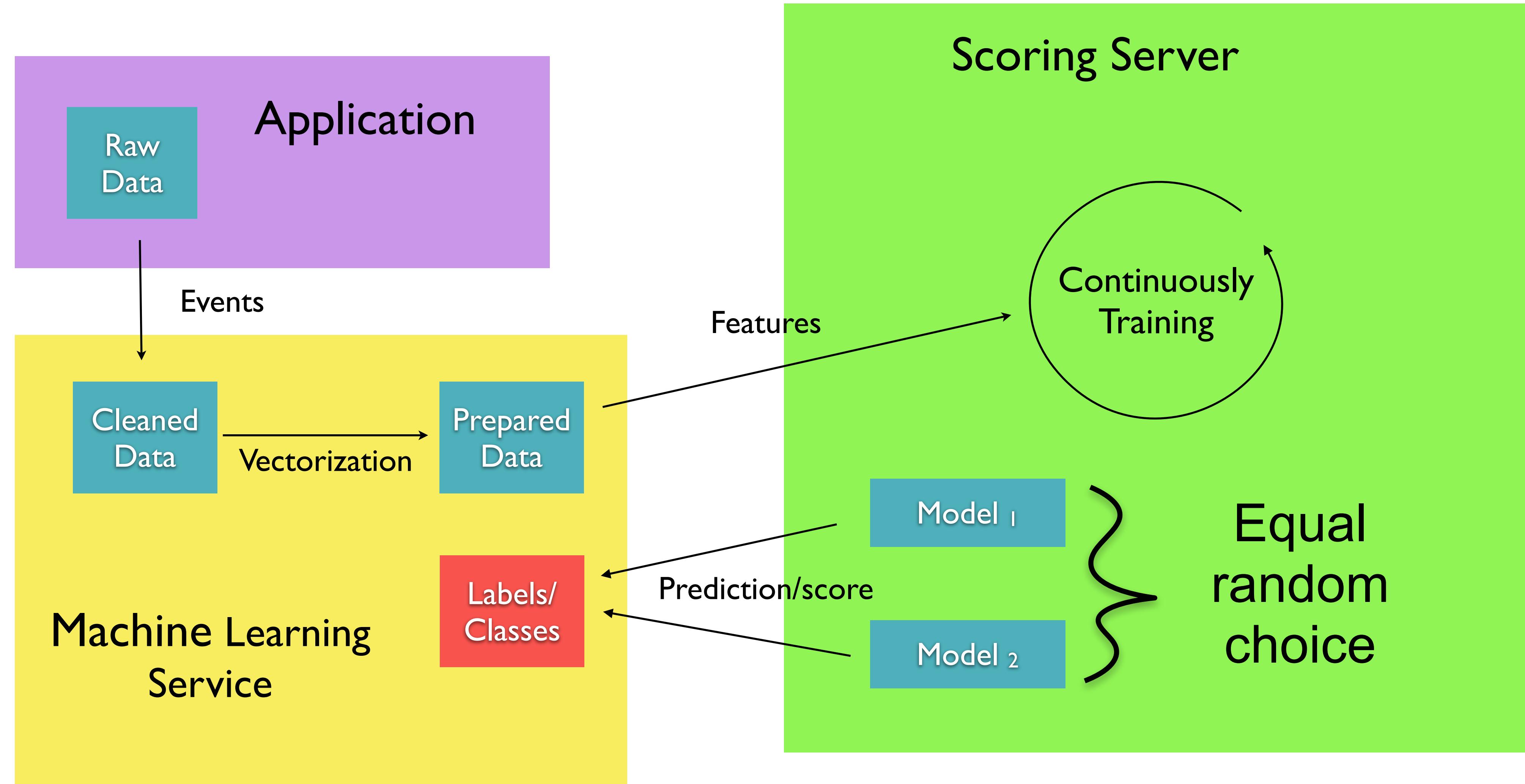


Online Evaluation Strategies

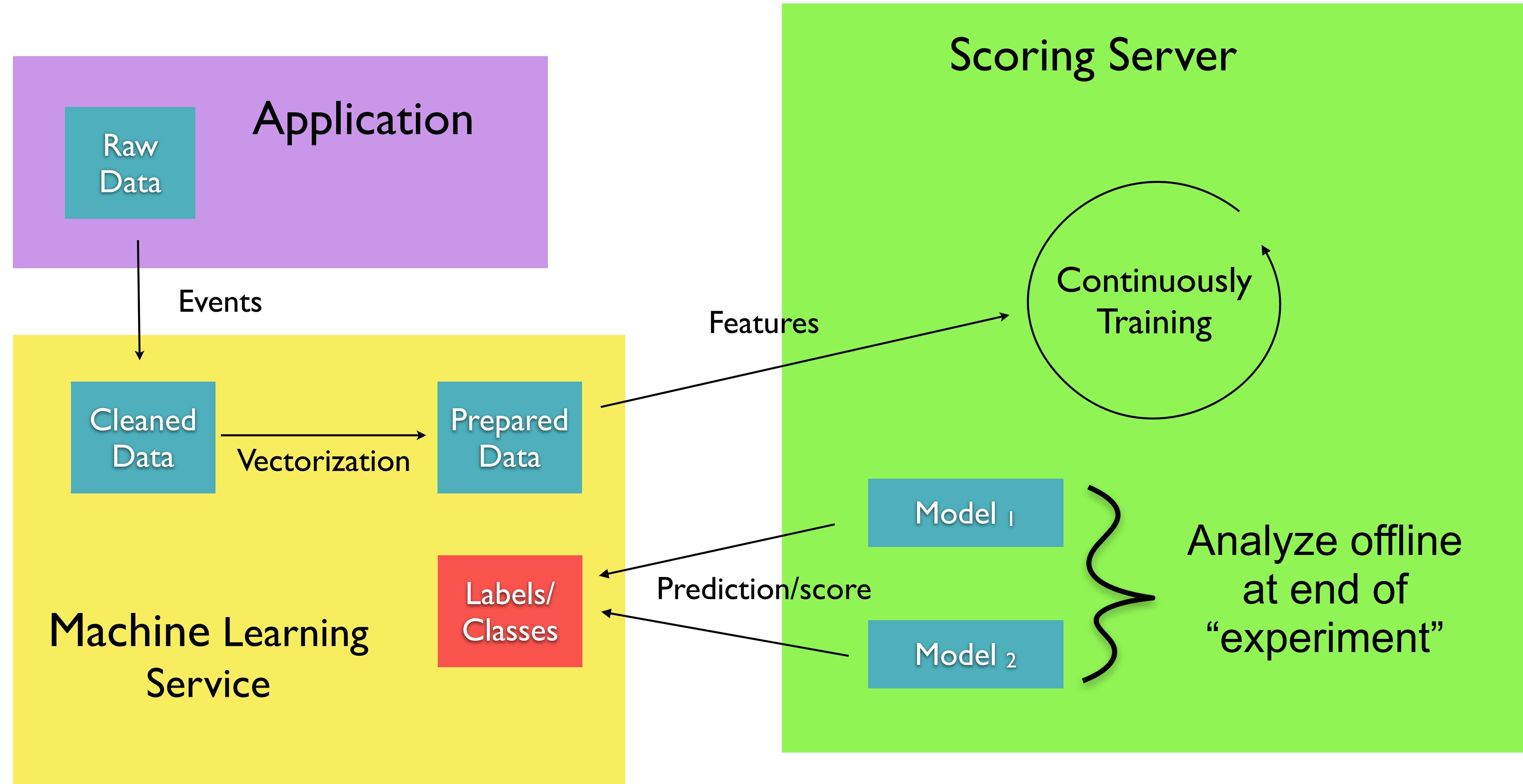
- Continuous Batch Offline Evaluation
- “Live” A/B Testing
- Multi-armed Bandit



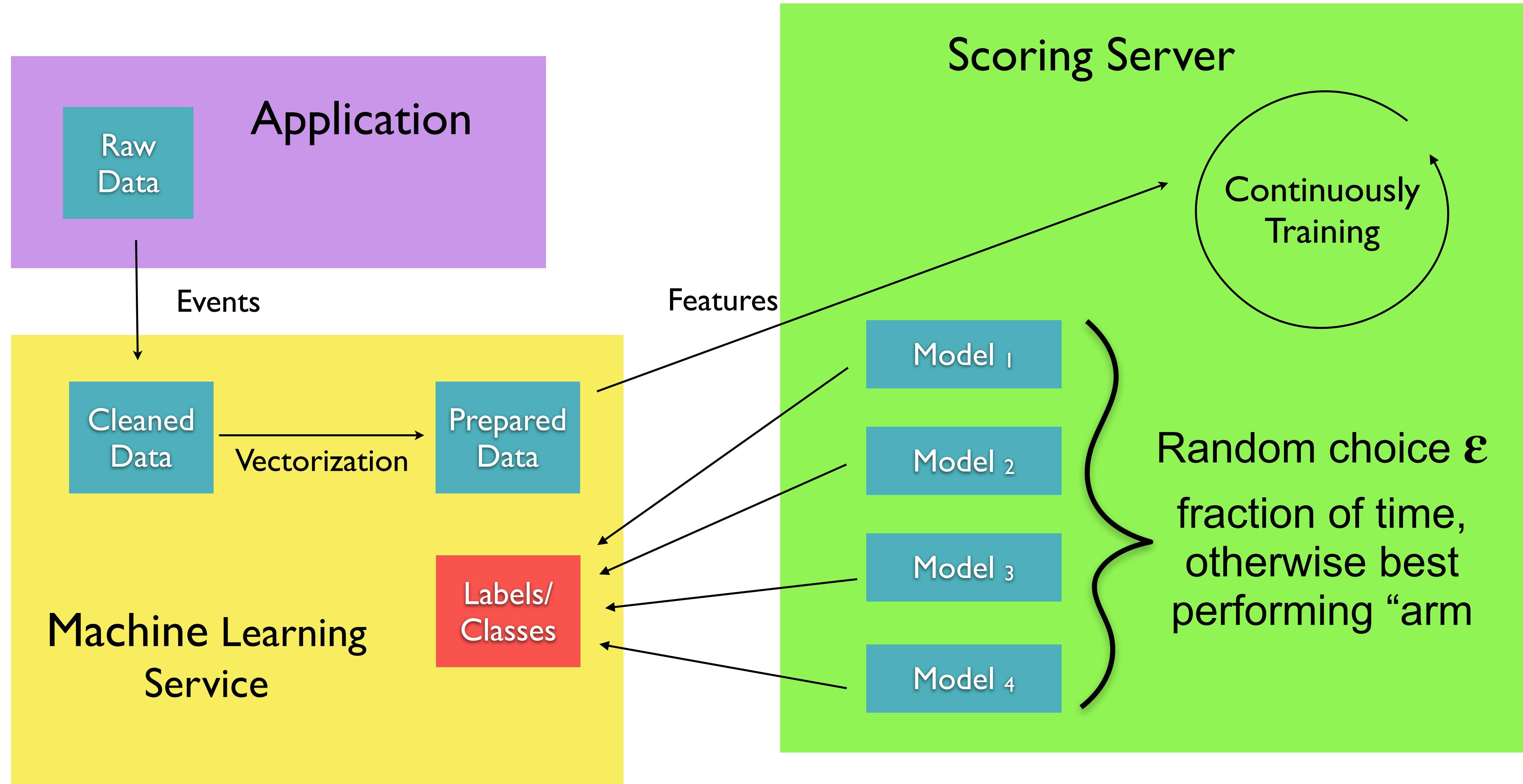
Live A/B Testing



Live A/B Testing



Multi-armed Bandit



Thank You 🙏

Materials — <https://jonathans.estate/scaling-data-science>

Jonathan — <https://jonathans.estate/youtube>

- Data science and deep learning tutorials