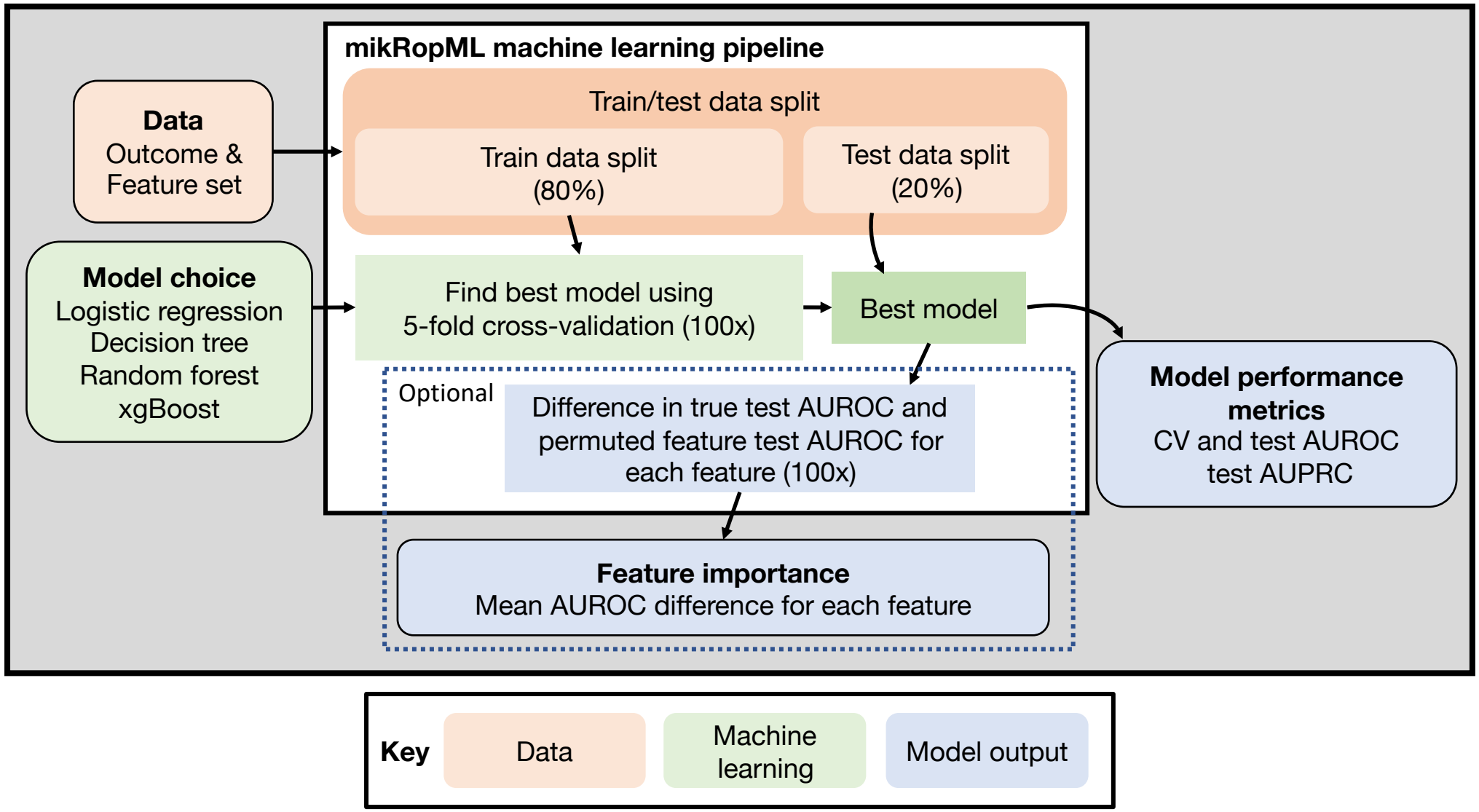
mikRopML machine learning pipeline
Data
Outcome &
Feature set
Train/test data split
Train data split
(80%)
Test data split
(20%)
Model choice
Logistic regression
Decision tree
Random forest
xgBoost
Find best model using
5-fold cross-validation (100x)
Best model
Optional
Difference in true test AUROC and
permuted feature test AUROC for
each feature (100x)
Model performance
metrics
CV and test AUROC
test AUPRC
Feature importance
Mean AUROC difference for each feature
Key
Data
Machine
learning
Model output