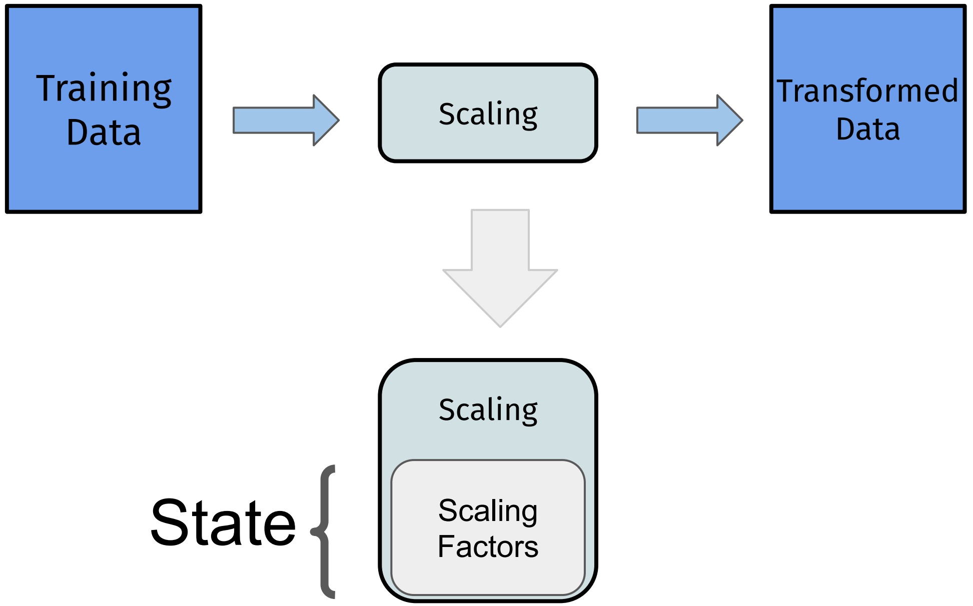
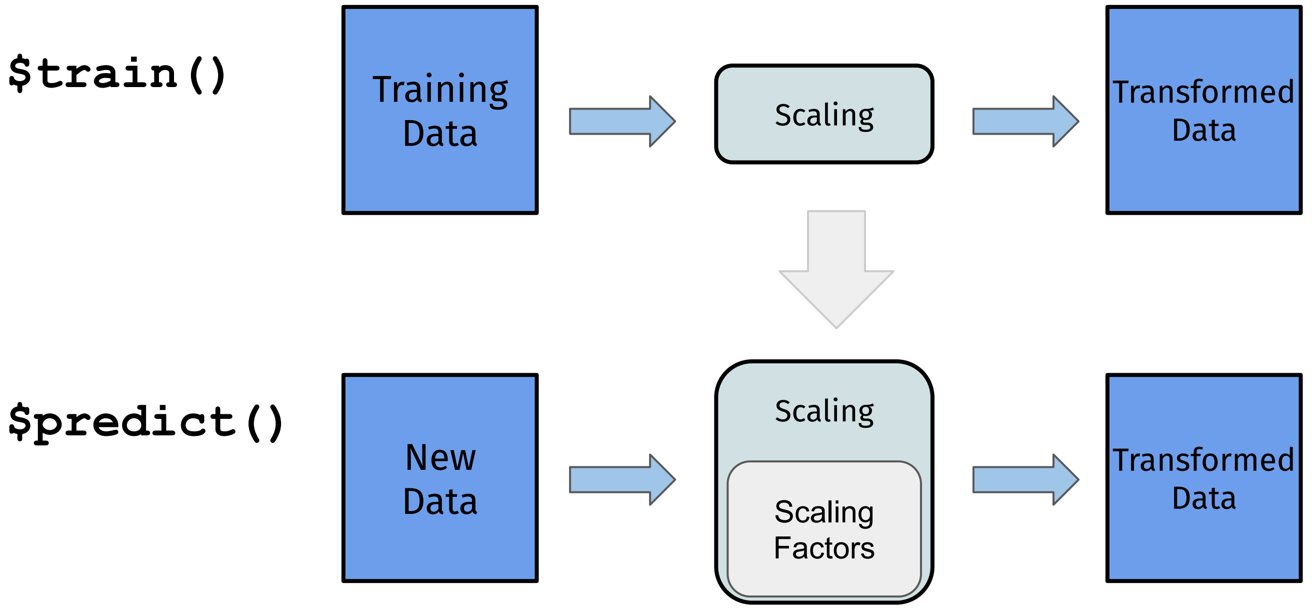
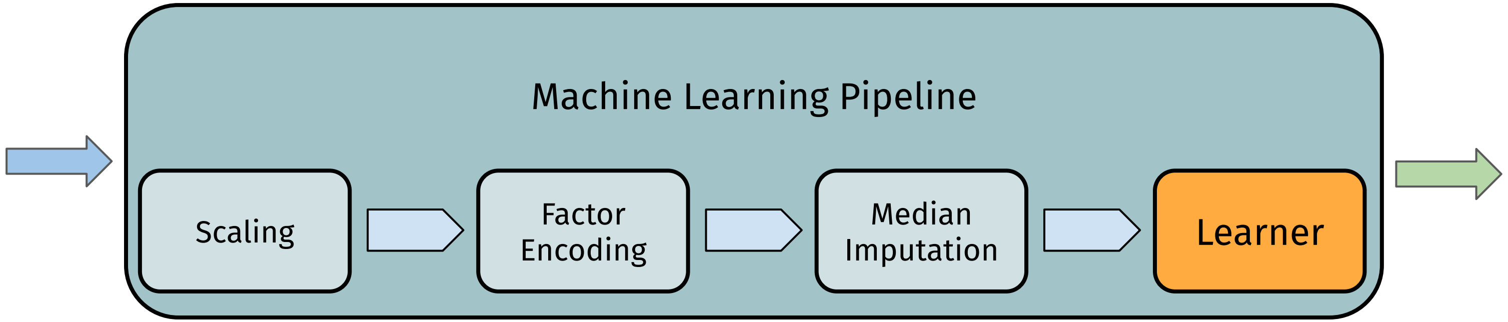
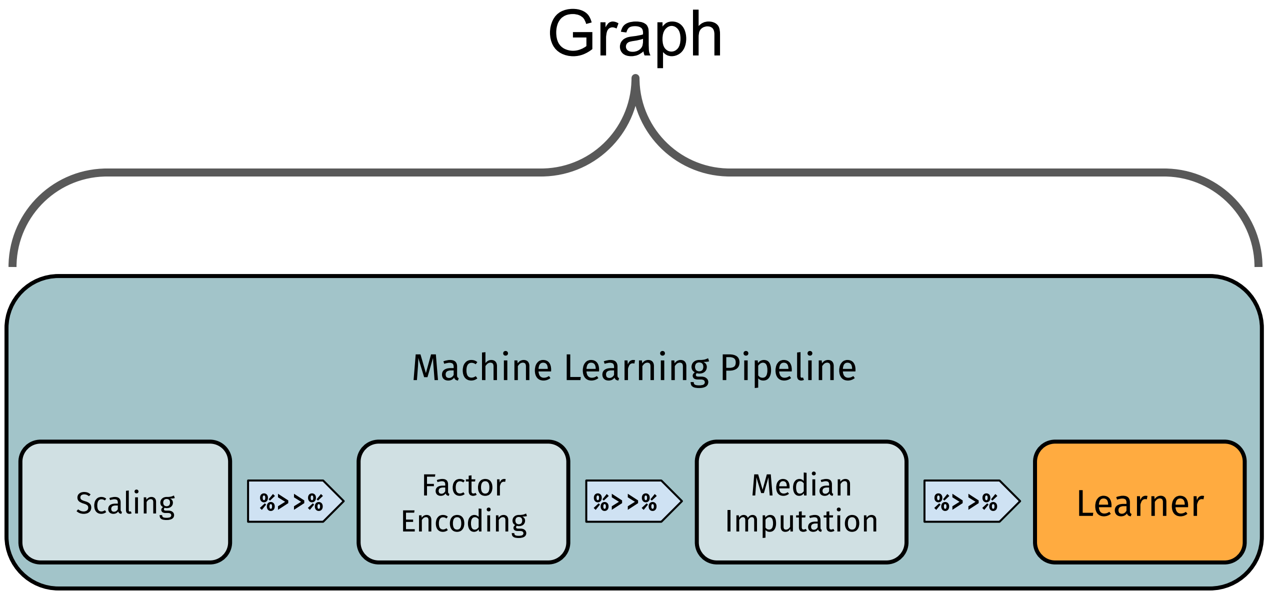
```
$train()
```

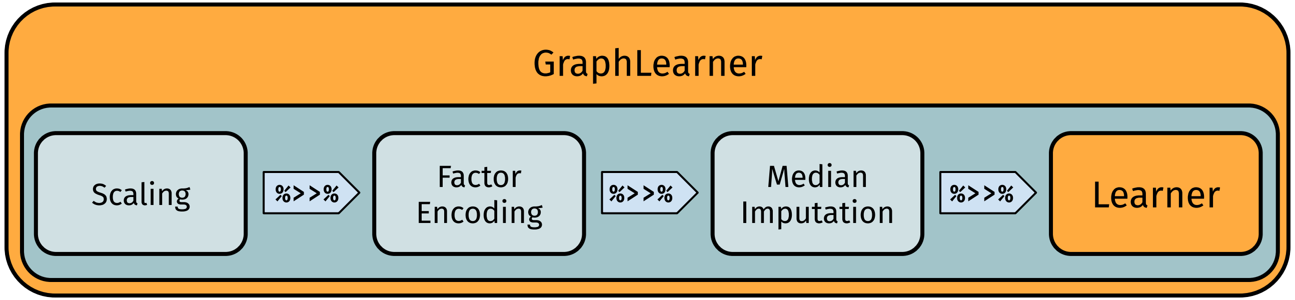

Training Data
Scaling
Transformed Data
Scaling
Scaling Factors

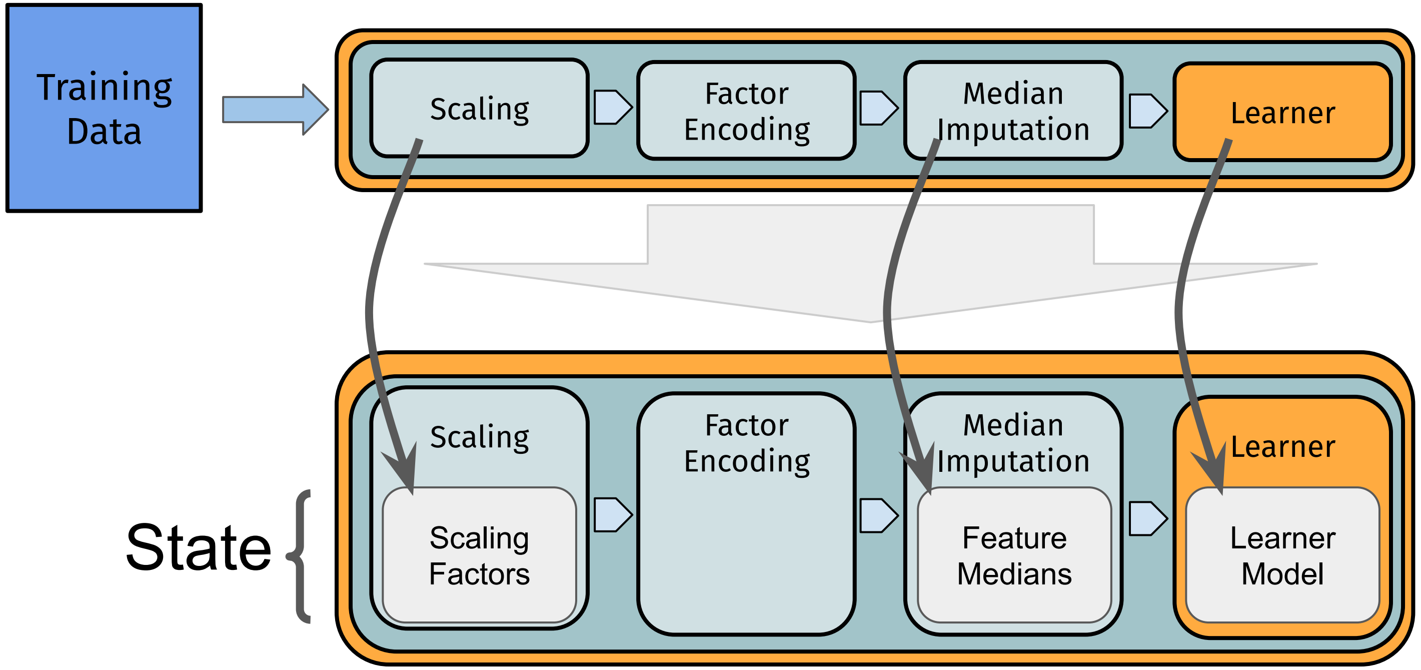
State

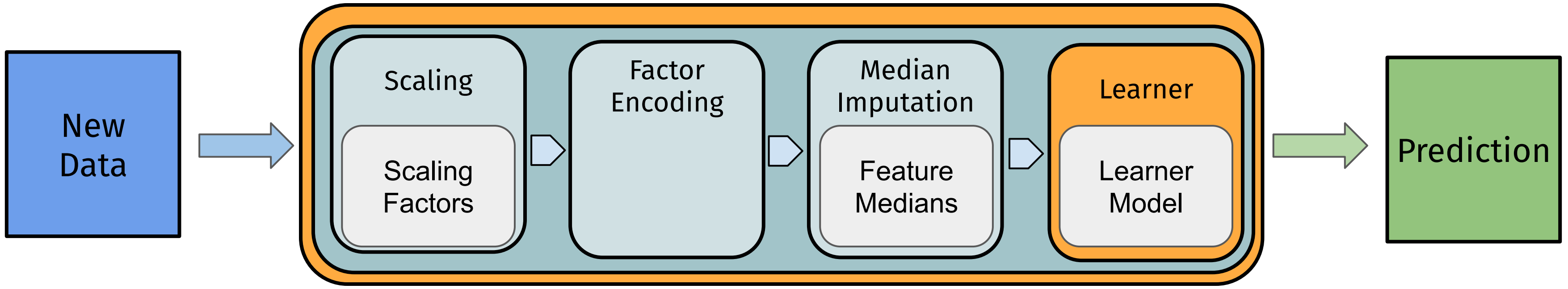

$train()
Training Data
Scaling
Transformed Data
$predict()
New Data
Scaling
Scaling Factors
Transformed Data

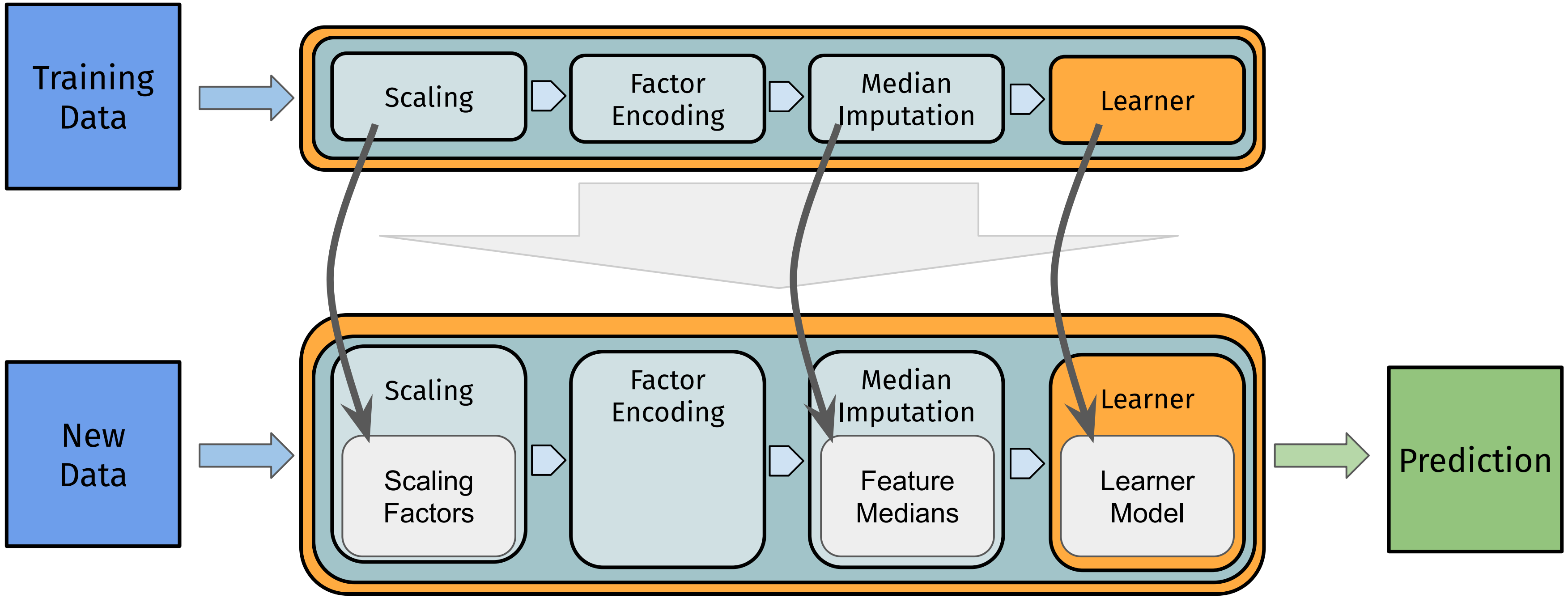

Machine Learning Pipeline
Scaling
Factor Encoding
Median Imputation
Learner

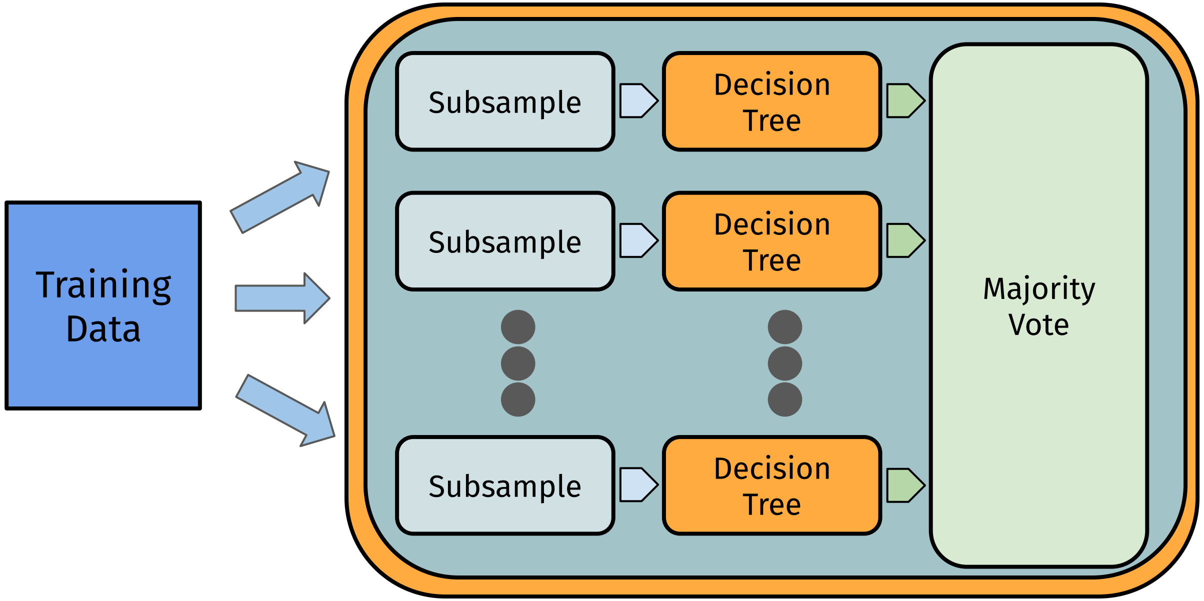

Graph
Machine Learning Pipeline
Scaling
%>>%
Factor Encoding
%>>%
Median Imputation
%>>%
Learner

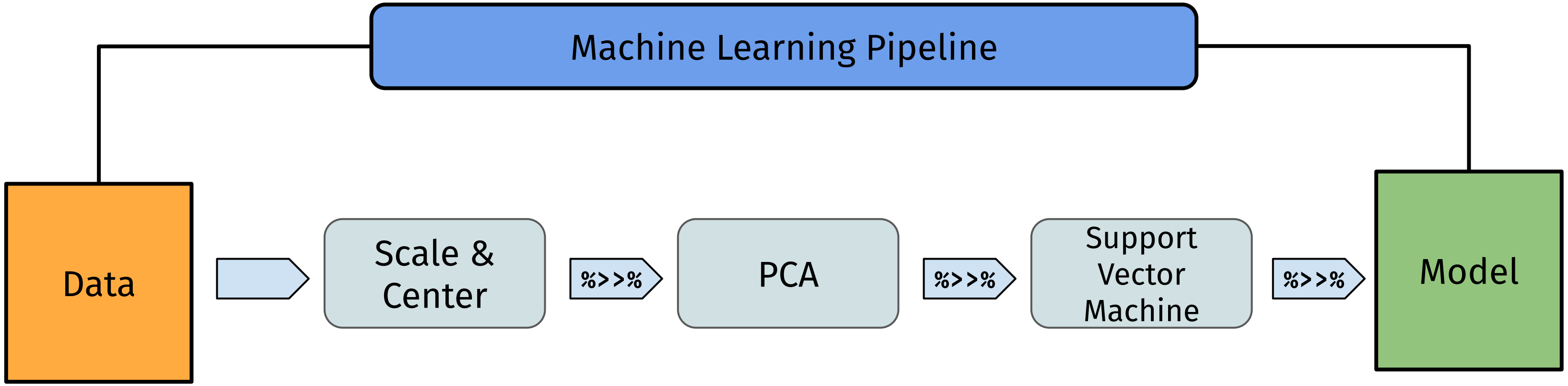

GraphLearner
Scaling
%>>%
Factor Encoding
%>>%
Median Imputation
%>>%
Learner


Training Data
Scaling
Factor Encoding
Median Imputation
Learner
Scaling
Factor Encoding
Median Imputation
Learner
State
Scaling Factors
Feature Medians
Learner Model

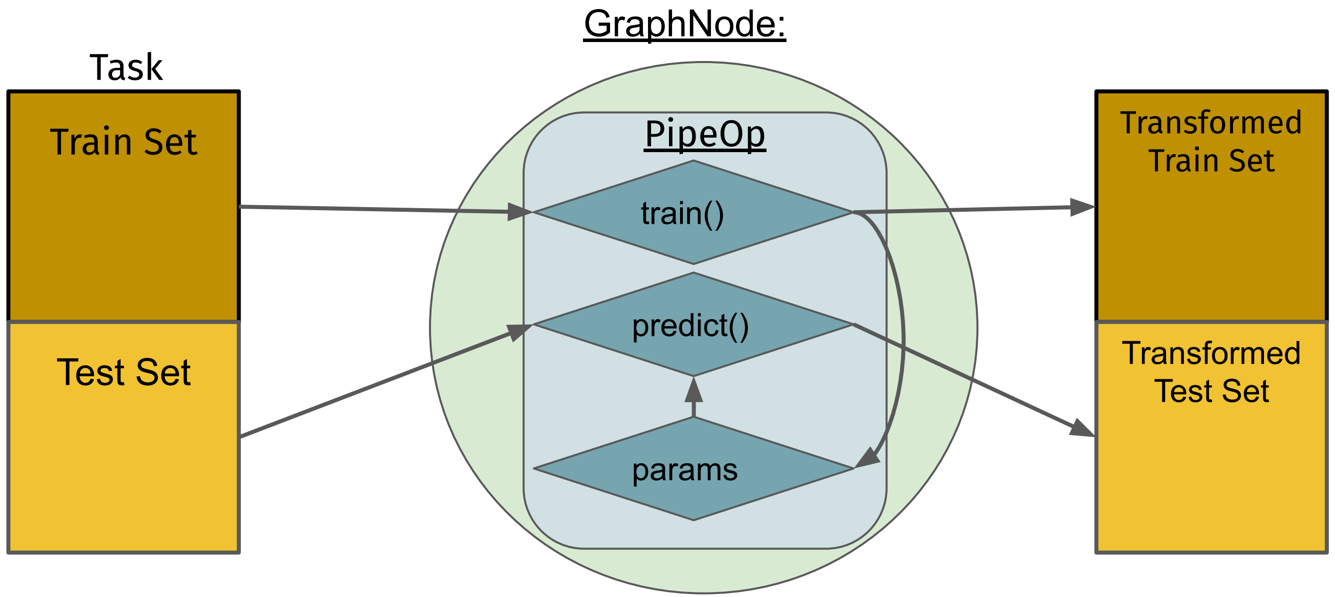

New Data
Scaling
Scaling Factors
Factor Encoding
Median Imputation
Feature Medians
Learner
Learner Model
Prediction


Training Data
Scaling
Factor Encoding
Median Imputation
Learner
New Data
Scaling
Scaling Factors
Factor Encoding
Median Imputation
Feature Medians
Learner
Learner Model
Prediction


Training Data
Subsample
Decision Tree
Subsample
Decision Tree
Subsample
Decision Tree
Majority Vote



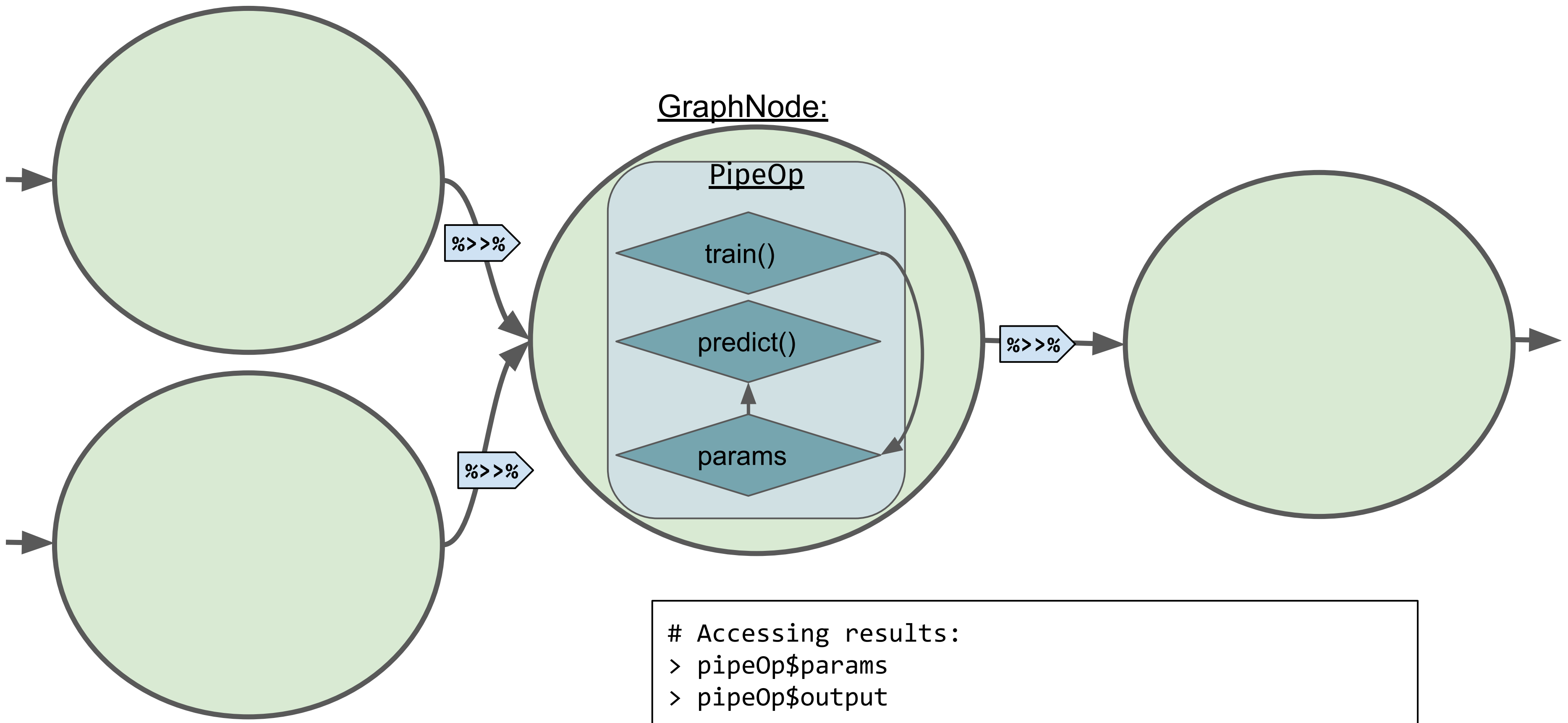
```
# Pseudo Code:
> pipeOpScale() %>>% pipeOpPCA() %>>% pipeOpLearner("svm")
```



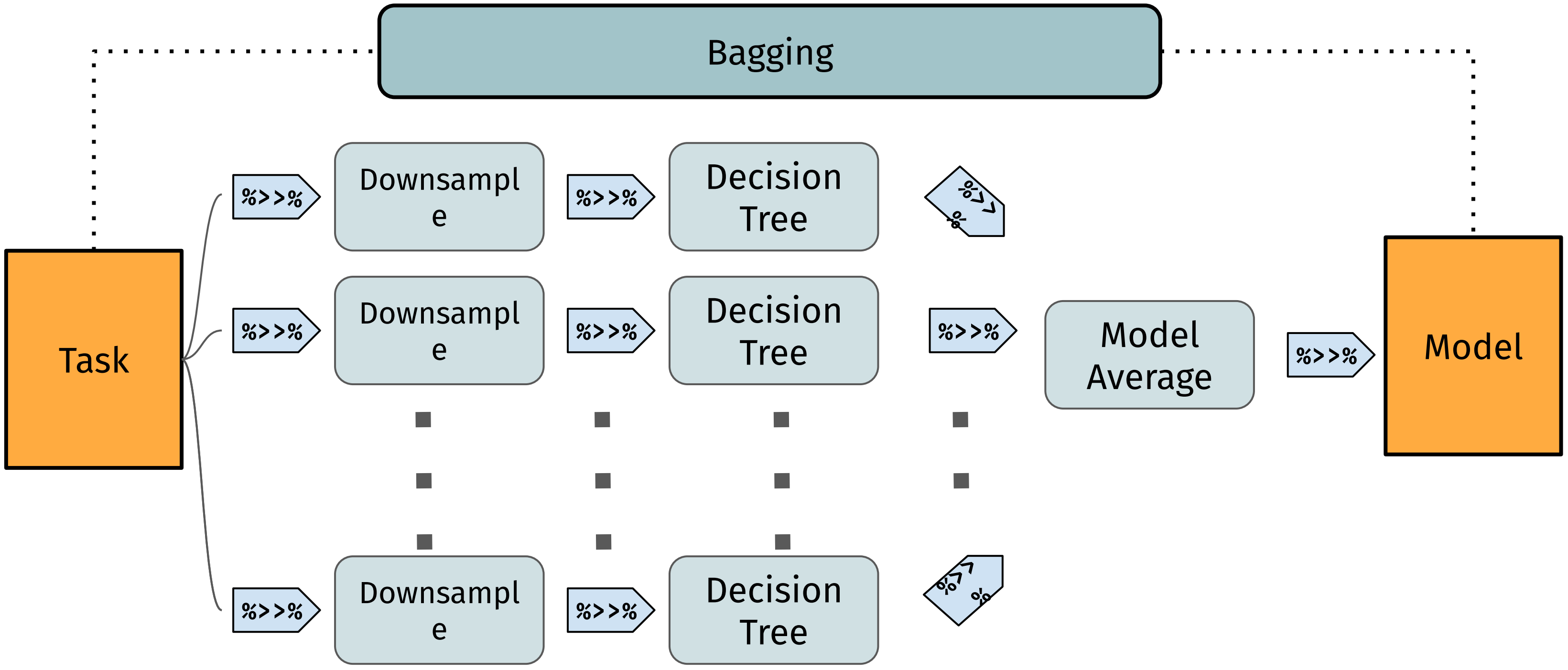
- `train()` saves transformation params and outputs transformed training data.

- `predict()` uses params and outputs transformed test data

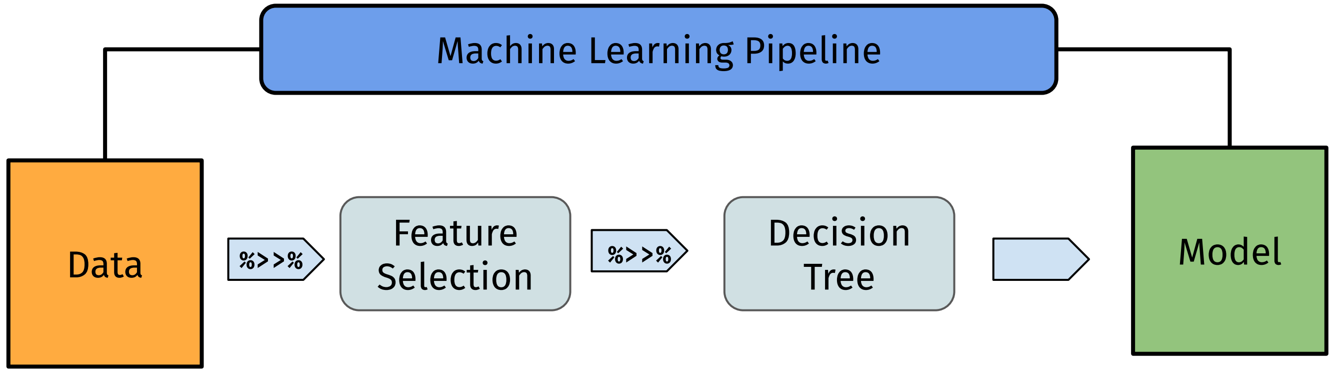
Multiple GraphNode's can be connected with " %>>% "



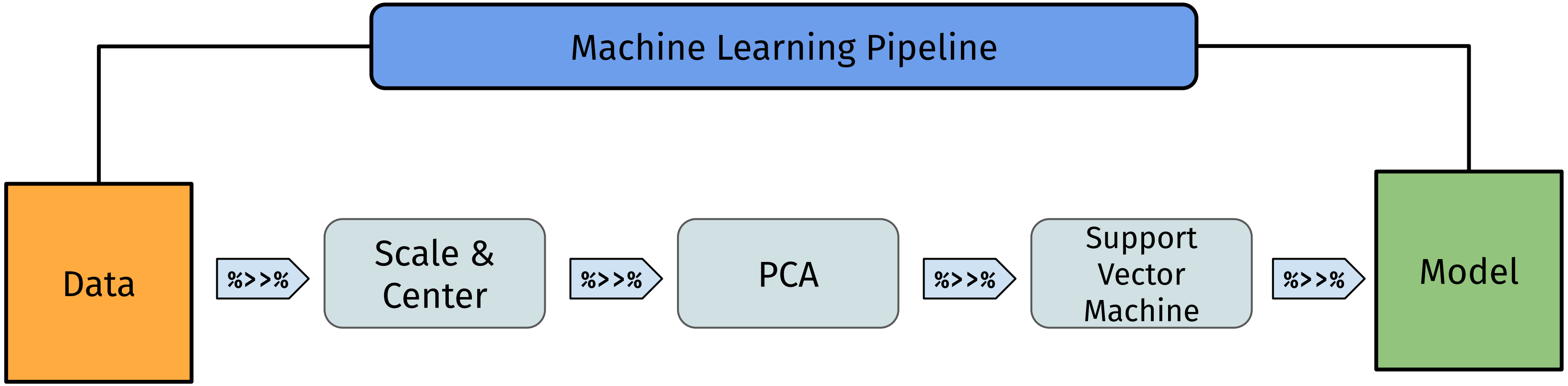
```
# Accessing results:
> pipeOp$params
> pipeOp$output
```



```
# Pseudo Code:
> rep(100, pipeOpDownsample() %>>% pipeOpLearner("rpart")) %>>% pipeOpModelAverage()
```


Machine Learning Pipeline
Data
%>>%
Feature Selection
%>>%
Decision Tree
Model

- <u>Pipelines provide:</u>
    - Multiple widely used operations
      (Scaling, PCA, Variable Selection, Imputation, Stacking and many others)
    - A clean, extendable interface for custom pipeline operators
    - A simple operator connection operator: **%>>%**
    - An abstraction for parallelization



```
# Pseudo Code:
> pipeOpScale() %>>% pipeOpPCA() %>>% pipeOpLearner(“svm”)
```