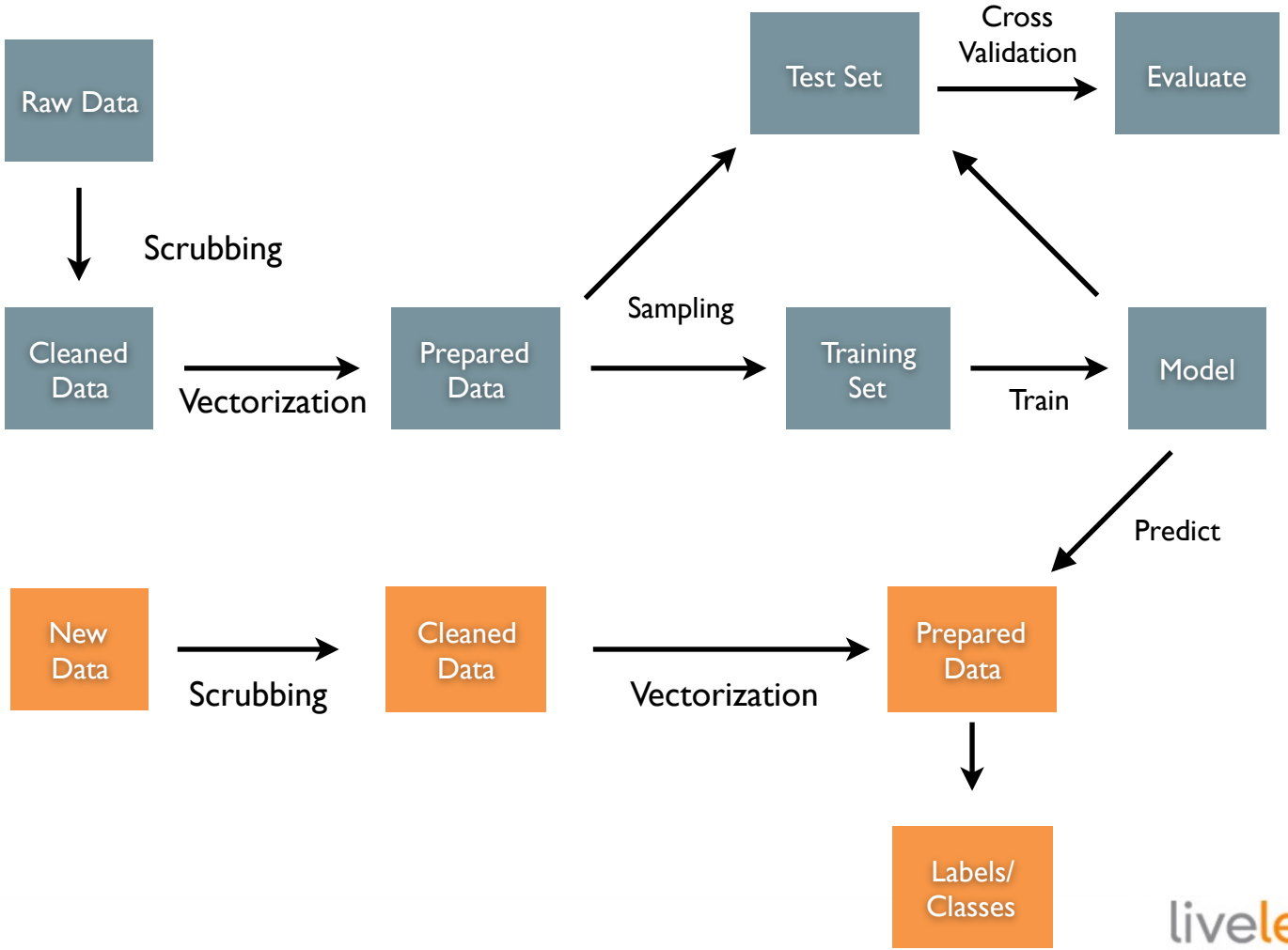
livelessons
video instruction from technology experts

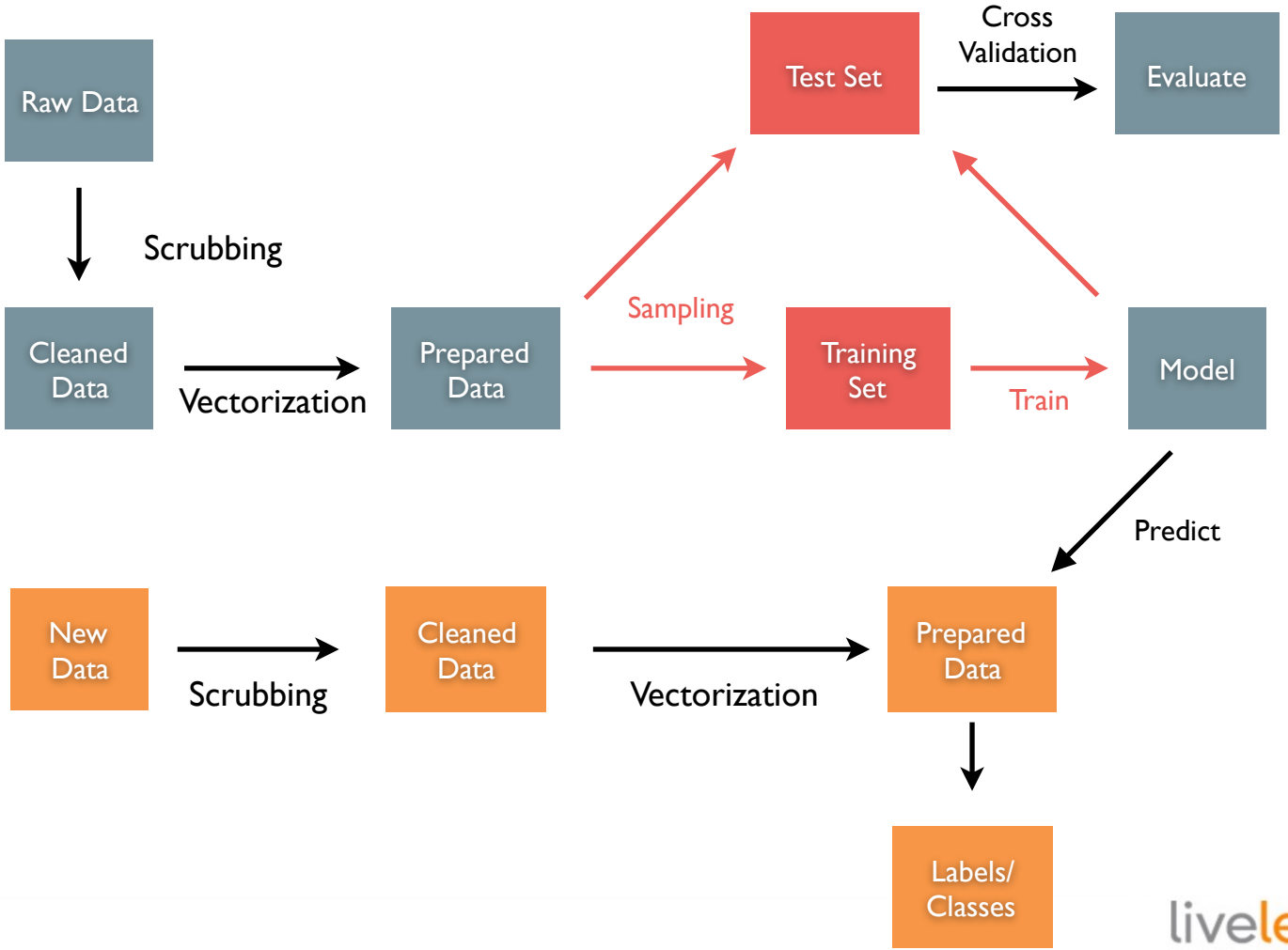
Lesson 5: Advanced Applications

5.4 Building a Model with MLlib

livelessons
©2016 Pearson, Inc.



livelessons
©2016 Pearson, Inc.



livelessons
©2016 Pearson, Inc.

Challenges

- “Very large dataset”
- Heterogeneous (continuous and discrete) noisy data
- Imbalanced classes
- (no context on data — anonymized)

livelessons
©2016 Pearson, Inc.

Review

- Built-in high performance ML algorithms for model building
- Utility functions for test-train split and evaluation
- Rapidly iterate on different models and hyper-parameters

livelessons
©2016 Pearson, Inc.

Next Up: Model Tuning and Evaluation

livelessons
©2016 Pearson, Inc.