

Lesson 5: Advanced Applications

5.2 The KDD Cup Competition: Preparing Data and Imputing Values

Review

- Two machine learning APIs: `spark.ml` and `MLlib`
- `MLlib` is a library of *à la carte* models and utility functions
- `spark.ml` is a higher level API for constructing ML *pipelines*
- In practical ML scenarios, many *more challenges* than simply *model optimization*.



Next Up: Machine Learning Models

