Apache Mahout

Apache Mahout is an Apache project to produce free implementations of distributed or otherwise scalable machine learning algorithms on the Hadoop platform. Mahout is a work in progress; the number of implemented algorithms has grown quickly but there are still various algorithms missing. While Mahout's core algorithms for clustering, classification and batch based collaborative filtering are implemented on top of Apache Hadoop using the map/reduce paradigm, it does not restrict contributions to Hadoop based implementations. Contributions that run on a single node or on a non-Hadoop cluster are welcomed. For example, the 'Taste' collaborative-filtering recommender component of Mahout was originally a separate project and can run stand-alone without Hadoop.

- Collaborative Filtering
- User and Item based recommenders
- K-Means, Fuzzy K-Means clustering
- Mean Shift clustering
- Dirichlet process clustering
- Latent Dirichlet Allocation
- Singular value decomposition
- Parallel Frequent Pattern mining
- Complementary Naive Bayes classifier
- Random forest decision tree based classifier