
H2O.ai

Supervised Learning

Statistical Analysis

- **Penalized Linear Models**: Super-fast, super-scalable, and interpretable
- **Naïve Bayes:** Straightforward linear classifier

Decision Tree Ensembles

- **Distributed Random Forest**: Easy-to-use tree-bagging ensembles
- **Gradient Boosting Machine**: Highly tunable tree-boosting ensembles
- **eXtreme Gradient Boosting**: Popular XGBoost algorithm in H2O

- **Deep neural networks**: Multi-layer feed-forward neural networks for standard data mining tasks
- **Convolutional neural networks:** Sophisticated architectures for pattern recognition in images, sound, and text

Unsupervised Learning

Clustering

- **K-means**: Partitions observations into similar groups; automatically detects number of groups

Dimensionality Reduction

- **Principal Component Analysis**: Transforms correlated variables to independent components
- **Generalized Low Rank Models:** Extends the idea of PCA to handle arbitrary data consisting of numerical, Boolean, categorical, and missing data

Anomaly Detection

Term Embeddings

- **Autoencoders**: Find outliers using a nonlinear dimensionality reduction technique
- **Word2vec:** Generate context-sensitive numerical representations of a large text corpus

Multilayer Perceptron

Deep Learning

Stacking

- **Stacked Ensemble**: Combine multiple types of models for better predictions

Aggregator

- **Aggregator:** Efficient, advanced sampling that creates smaller data sets from larger data sets

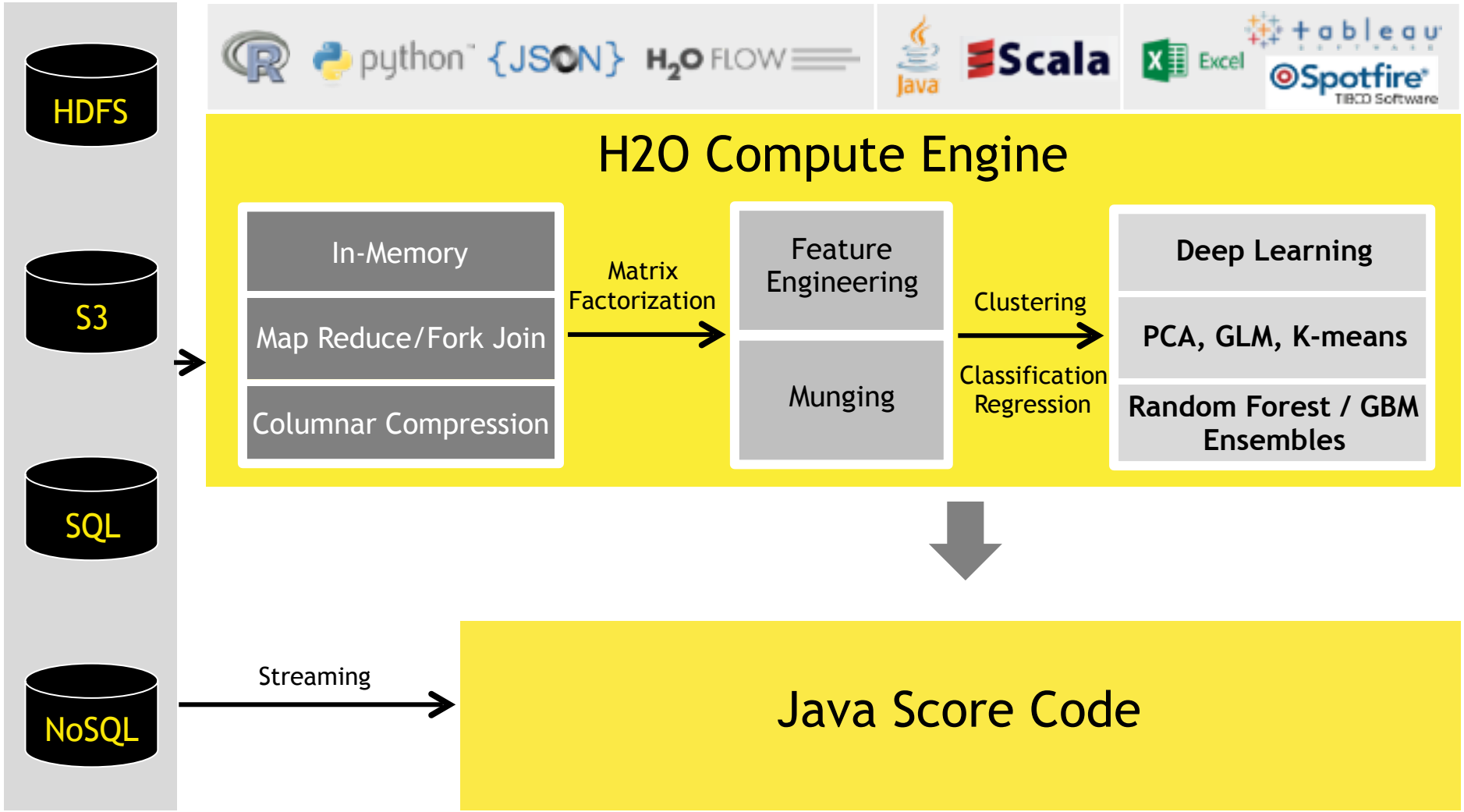
H2O Machine Learning Methods

Neural Networks

AutoML

- **Automatic Machine Learning**: Automated exploration of supervised learning approaches

Data and Client Agnostic



H2O Machine Learning Methods

Supervised Learning

Statistical Analysis

- **Penalized Linear Models**: Super-fast, super-scalable, and interpretable
- **Naïve Bayes:** Straightforward linear classifier

Decision Tree Ensembles

- **Distributed Random Forest**: Easy-to-use tree-bagging ensembles
- **Gradient Boosting Machine**: Highly tunable tree-boosting ensembles
- **eXtreme Gradient Boosting**: Popular XGBoost algorithm in H2O

Stacking

- **Stacked Ensemble**: Combine multiple types of models for better predictions

AutoML

- **Automatic Machine Learning**: Automated exploration of supervised learning approaches

Unsupervised Learning

Clustering

- **K-means**: Partitions observations into similar groups; automatically detects number of groups

Dimensionality Reduction

- **Principal Component Analysis**: Transforms correlated variables to independent components
- **Generalized Low Rank Models:** Extends the idea of PCA to handle arbitrary data consisting of numerical, Boolean, categorical, and missing data

Aggregator

- **Aggregator:** Efficient, advanced sampling that creates smaller data sets from larger data sets

Neural Networks

Multilayer Perceptron

Deep Learning

- **Deep neural networks**: Multi-layer feed-forward neural networks for standard data mining tasks
- **Convolutional neural networks:** Sophisticated architectures for pattern recognition in images, sound, and text

Anomaly Detection

Term Embeddings

- **Autoencoders**: Find outliers using a nonlinear dimensionality reduction technique
- **Word2vec:** Generate context-sensitive numerical representations of a large text corpus