

Treasure Machine Learning Function Reference

Treasure Machine Learning (ML) is based on Apache Hivemall, which is a scalable machine learning library that runs on Apache Hive. Hivemall is designed to be scalable to the number of training instances and the number of training features. For more information, see the [Hivemall User Guide](#).

- [Supported Algorithms](#)
 - [Classification](#)
 - [Regression](#)
 - [Recommendation](#)
 - [k-Nearest Neighbor](#)
 - [Feature Engineering](#)
- [Hivemall Evaluation UDFs](#)
 - [Binary Classification Metrics](#)
 - [Ranking Measures](#)

Supported Algorithms

Hivemall provides machine learning functionality and feature engineering functions through the Hive user-defined functions (UDFs), user-defined aggregation function (UDAFs), and user-defined tabular functions (UDTFs).

Classification

- Perceptron
- Passive Aggressive (PA, PA1, PA2)
- Confidence Weighted (CW)
- Adaptive Regularization of Weight Vectors (AROW)
- Soft Confidence Weighted (SCW1, SCW2)
- AdaGradRDA (with hinge loss)
- RandomForest
 - Example: [Multiclass classification on Iris Dataset](#)
- Factorization Machines

Regression

- Logistic Regression using Stochastic Gradient Descent
- AdaGrad / AdaDelta (with logistic loss)
- Passive Aggressive Regression (PA1, PA2)
- AROW regression
- RandomForest
- Factorization Machines
 - Example: [Movielens rating prediction](#)

Recommendation

- Matrix Factorization (sgd, adagrad)
 - Example: [Movielens rating prediction](#)
- Minhash (LSH with jaccard index)

k-Nearest Neighbor

- Minhash (LSH with jaccard index)
- b-Bit minhash
- Brute-force search using Cosine similarity

Feature Engineering

- Feature hashing (MurmurHash, SHA1)
- Feature scaling (Min-Max Normalization, Z-Score)
- Feature instances amplifier that reduces iterations on training
- [TF-IDF vectorizer](#)

Hivemall Evaluation UDFs

Evaluation UDFs are useful for evaluating the accuracy of your machine learning model.

Binary Classification Metrics

Signature

```
auc(probability, truth_label)
fmeasure(truth_label, predicted_label)
```

Description

See the [Hivemall](#) user guide.

Ranking Measures

Signature

```
auc(recommend_list, truth_list, k)
precision_at(recommend_list, truth_list, k)
recall_at(recommend_list, truth_list, k)
mrr(recommend_list, truth_list, k)
average_precision(recommend_list, truth_list, k)
hitrate(recommend_list, truth_list, k)
```

Description

See the [Hivemall user guide](#).