Thank you for accepting the rules.

Higgs Boson Machine Learning Challenge

Monday, May 12, 2014 — 2 months to go — $13,000 • 713 teams — Monday, September 15, 2014

Dashboard

Home
- Data
- Make a submission

Information
- Description
- Evaluation
- Rules
- Prizes
- About the Sponsors
- Timeline

Forum
Leaderboard
My Team
My Submissions

Forum (37 topics)

R's gbm vs. Python's xgboost
47 minutes ago

A quick recipe with scikit-learn
2 days ago

Technical documentation
3 days ago

Rank order for background- and signal-like events
4 days ago

Submission HiggsBoson failing.
6 days ago

Derived features
7 days ago

Competition Details » Get the Data » Make a submission

Data Files

File Name	Available Formats
random_submission	.zip (2.58 mb)
training	.zip (16.89 mb)
test	.zip (34.79 mb)
HiggsBosonCompetition_AMSMetric_rev1	.py (3.15 kb)

File descriptions

- **training.csv** - Training set of 250000 events, with an ID column, 30 feature columns, a weight column and a label column.
- **test.csv** - Test set of 550000 events with an ID column and 30 feature columns.
- **random_submission** - Sample submission file in the correct format. File format is described on the Evaluation page.
- **HiggsBosonCompetition_AMSMetric** - Python script to calculate the competition evaluation metric.

For detailed information on the semantics of the features, labels, and weights, see the technical documentation from the LAL website on the task.

Some details to get started:

- all variables are floating point, except **PRI_jet_num** which is integer
- variables prefixed with **PRI** (for PRImitives) are “raw” quantities about the bunch collision as measured by the detector.
- variables prefixed with **DER** (for DERived) are quantities computed from the primitive features, which were selected by the physicists of ATLAS
- it can happen that for some entries some variables are meaningless or

cannot be computed; in this case, their value is –999.0, which is outside the normal range of all variables