kaggle   Search kaggle   Q   **Competitions   Datasets   Kernels   Discussion   Jobs**

Blue Book for Bulldozers

Predict the auction sale price for a piece of heavy equipment to create a "blue book" for bulldozers.
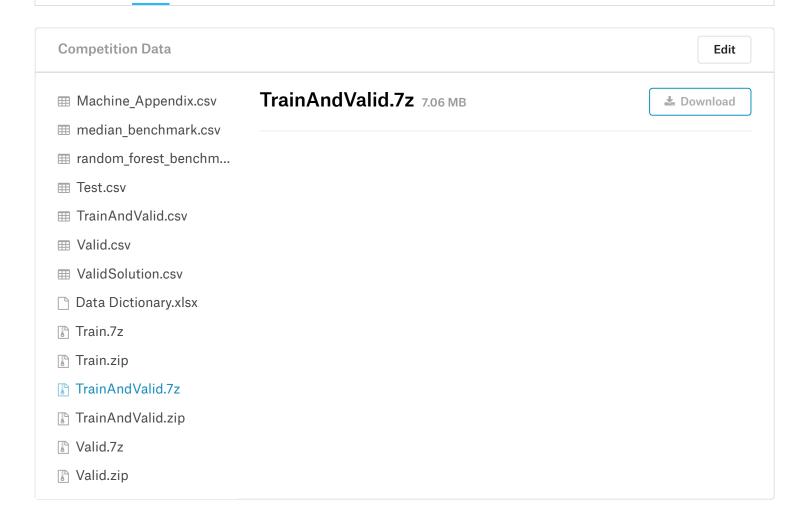
$10,000 · 475 teams · 5 years ago

Overview   **Data**   Discussion   Leaderboard   Rules

Competition Data   Edit

| ▦ Machine_Appendix.csv |
| ▦ median_benchmark.csv |
| ▦ random_forest_benchm... |
| ▦ Test.csv |
| ▦ TrainAndValid.csv |
| ▦ Valid.csv |
| ▦ ValidSolution.csv |
| ▯ Data Dictionary.xlsx |
| ▤ Train.7z |
| ▤ Train.zip |
| ▤ TrainAndValid.7z |
| ▤ TrainAndValid.zip |
| ▤ Valid.7z |
| ▤ Valid.zip |

TrainAndValid.7z 7.06 MB   ⬇ Download

Data Description

View and download the benchmark code from Github

For this competition, you are **predicting the sale price of bulldozers sold at auctions**.

The data for this competition is split into three parts:

- **Train.csv** is the training set, which contains data through the end of 2011.

- **Valid.csv** is the validation set, which contains data from January 1, 2012 - April 30, 2012 You make predictions on this set throughout the majority of the competition. Your score on this set is used to create the public leaderboard.

- **Test.csv** is the test set, which won't be released until the last week of the competition. It contains data from May 1, 2012 - November 2012. Your score on the test set determines your final rank for the competition.

The key fields are in train.csv are:

- SalesID: the uniue identifier of the sale

- MachineID: the unique identifier of a machine. A machine can be sold multiple times

- saleprice: what the machine sold for at auction (only provided in train.csv)

- saledate: the date of the sale

There are several fields towards the end of the file on the different options a machine can have. The descriptions all start with "machine configuration" in the data dictionary. Some product types do not have a particular option, so all the records for that option variable will be null for that product type. Also, some sources do not provide good option and/or hours data.

The machine_appendix.csv file contains the correct year manufactured for a given machine along with the make, model, and product class details. There is one machine id for every machine in all the competition datasets (training, evaluation, etc.).

Our Team Terms Privacy Contact/Support