

Sparkling Water

MICHAL MALOHLAVA

ALEX TELLEZ

JESSICA LANFORD

<http://h2o.gitbooks.io/sparkling-water-and-h2o/>

August 2015: First Edition

Sparkling Water
by Michal Malohlava, Alex Tellez & Jessica Lanford

Published by H2O.ai, Inc.
2307 Leghorn St.
Mountain View, CA 94043

©2015 H2O.ai, Inc. All Rights Reserved.

August 2015: First Edition

Photos by ©H2O.ai, Inc.

While every precaution has been taken in the preparation of this book, the publisher and authors assume no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein.

Printed in the United States of America.

Contents

1	What is H2O?	3
2	Introduction	4
3	Appendix A: Complete parameter list	5
4	Appendix B: References	6

1 What is H2O?

H2O is fast, scalable, open-source machine learning and deep learning for Smarter Applications. With H2O, enterprises like PayPal, Nielsen, Cisco, and others can use all their data without sampling to get accurate predictions faster. Advanced algorithms, like Deep Learning, Boosting, and Bagging Ensembles are built-in to help application designers create smarter applications through elegant APIs. Some of our initial customers have built powerful domain-specific predictive engines for Recommendations, Customer Churn, Propensity to Buy, Dynamic Pricing, and Fraud Detection for the Insurance, Healthcare, Telecommunications, AdTech, Retail, and Payment Systems industries.

Using in-memory compression, H2O handles billions of data rows in-memory, even with a small cluster. To make it easier for non-engineers to create complete analytic workflows, H2O's platform includes interfaces for R, Python, Scala, Java, JSON, and Coffeescript/JavaScript, as well as a built-in web interface, Flow. H2O was built alongside (and on top of) Hadoop and Spark Clusters and typically deploys within minutes.

H2O includes many common machine learning algorithms, such as generalized linear modeling (linear regression, logistic regression, etc.), Naïve Bayes, principal components analysis, time series, k-means clustering, and others. H2O also implements best-in-class algorithms at scale, such as Random Forest, Gradient Boosting and Deep Learning. Customers can build thousands of models and compare the results to get the best predictions.

H2O is nurturing a grassroots movement of physicists, mathematicians, and computer scientists to herald the new wave of discovery with data science by collaborating closely with academic researchers and Industrial data scientists. Stanford university giants Stephen Boyd, Trevor Hastie, Rob Tibshirani advise the H2O team on building scalable machine learning algorithms. With hundreds of meetups over the past two years, H2O has become a word-of-mouth phenomenon, growing amongst the data community by a hundred-fold, and is now used by 12,000+ users and is deployed using R, Python, Hadoop, and Spark in 2000+ corporations.

Try it out

H2O's R package can be installed from CRAN at <https://cran.r-project.org/web/packages/h2o/>. A Python package can be installed from PyPI at <https://pypi.python.org/pypi/h2o/>. Download H2O directly from <http://h2o.ai/download>.

Join the community

Visit the open source community forum at <https://groups.google.com/d/forum/h2ostream>. To learn about our meetups, training sessions, hackathons, and product updates, visit <http://h2o.ai>.

2 Introduction

3 Appendix A: Complete parameter list

4 Appendix B: References