

Machine learning on H2O.ai using R language

H2O is a distributed in-memory machine learning platform with linear scalability and predictive analytics platform that is entirely open source. The most popular statistical and machine learning algorithms, such as deep learning, generalised linear models, and gradient boosted machines, are supported by H2O. A scoreboard of the top models is generated by H2O's industry-leading AutoML feature, which automatically evaluates all algorithms and their hyperparameters.

H2O.ai is focused on bringing AI to businesses through software. 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 JavaScript, as well as a built-in web interface, Flow. H2O platform is extremely popular in both the R & Python communities.

H2O divides the dataset into clunks and runs separately on each cluster and at the end collects that data and shows it as final output on a single cluster, that is how it uses the scalability and distributed feature.

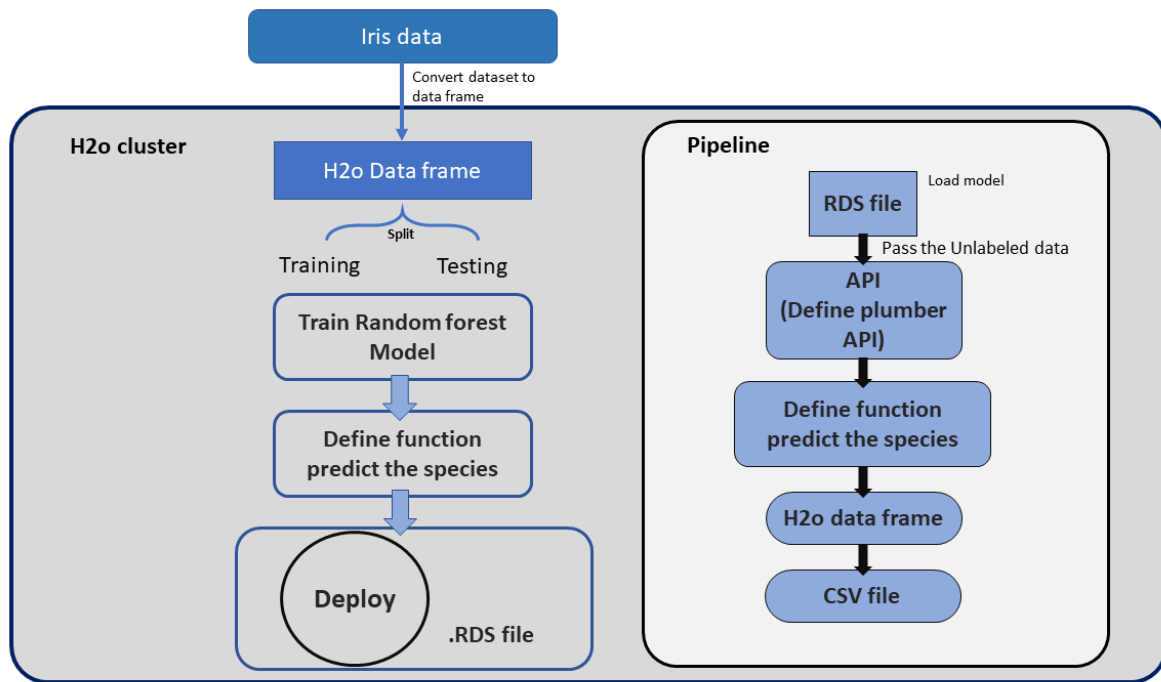
H2O is used because of its Scalability, Leading algorithm, Access from R, Python, Flow and more ..., AutoML, Distributed, In-Memory Processing Simple Deployment features.

So, In this use case we used H2O.ai using R language.

What is h2o driverless AI?

H2O Driverless AI (H2O.ai) is an artificial intelligence (AI) platform for automatic machine learning. Driverless AI automates some of the most difficult data science and machine learning workflows, such as feature engineering, model validation, model tuning, model selection, and model deployment. Model deployment could be easier here.

The iris dataset, which includes the three species Setosa, Versicolor and Virginica are used to develop the machine learning model in this use case. Although these species appear similar, they have different traits because some varieties of the iris flower are poisonous, the medical



industry occasionally needs to identify the species in order to manufacture skin products. Therefore, we require this kind of model to avoid those poisonous flower species and identify the correct species. To implement this flower species detection Random Forest Model is used from 'h2o' library in R language, for that following steps are followed:

First import h2o library in R then connect R with h2o using `h2o.init()` to start the cluster.

Fit the random forest model using training, testing and validation on that and make a function to predict the values and then deploy that model using h2o And H2O will speed up the processing by dividing the dataset into smaller portions and distributes these portions across multiple machines in cluster each machine in the cluster processes a subset of data and send result to h2o cluster and then h2o aggregates the result into final output.