

In recent years, there has been a tremendous effort from different technology companies trying to create data science platforms to streamline machine learning and data science operationalization. For instance: [KNIME](#), [H2O](#), [Databricks](#) etc. I have decided to test out H2O, one of the leading machine learning platform.

```
{% include figure.liquid loading="eager" path="assets/img/h2o/h2oologo.png" title="example image"
class="img-fluid rounded z-depth-1" %}
```

Installation and Getting Started

Installation is easy following [the download instructions](#). H2O flow can be opened within localhost, providing a notebook-style (kind of like Jupyter Notebook) for performing data science.

```
{% include figure.liquid loading="eager" path="assets/img/h2o/h2oflow.png" title="example image"
class="img-fluid rounded z-depth-1" %}
```

The UI is simple to follow and does not require any prior programming experience to use. Most of the data preparation procedures are point and click, and are very intuitive.

Model Training

AutoML is also very straight forward in this case.

```
{% include figure.liquid loading="eager" path="assets/img/h2o/h2oautoml.png" title="example image"
class="img-fluid rounded z-depth-1" %}
```

I was able to input the raw data from [this competition](#) without any preprocessing such as performing one-hot-encoding etc. The data contains both numerical values and string values. AutoML takes care of all this and was able to perform predictions. I have chosen not to exclude any models from being excluded from training and set the training time to be 1 hour.

Results

I uploaded the prediction results onto Kaggle and received a decent score. This prediction outperforms my previous attempt using AutoML Tables on GCP and places me in the upper 46%.

```
{% include figure.liquid loading="eager" path="assets/img/h2o/h2okagglersresults.png" title="example
image" class="img-fluid rounded z-depth-1" %}
```

Summary

Overall H2O is very simple to pick up and can be used by anyone without programming experience but a strong interest in data science. The entire processing from data ingestion to prediction only took 1.5 hours (including installation), with 1 hour spent on training. The prediction also outperforms AutoML Tables services provided on GCP. I highly recommend trying out H2O.