AutoML - Primeras Impresiones con H2O

# Introduccion

En el presente trabajo de investigación utilizamos 3 notebooks para comparar la manera de trabajar con el *approach* tradicional de desarrollo de modelos de machine learning contra las características de utilizar herramientas de automL como H20.

Utilizamos dos notebooks siguiendo el modelo tradicional para entrenar un algoritmo del estilo Random Forest y una red nuronal convolucional sobre el dataset MNIST, ambos con buenos resultados de scoring.

Por último utilizamos la potencia de la característica que provee H2O para correr diferentes algoritmos de ML (en nuestro caso nos enfocamos a los algoritmos de DeepLearning) de manera automática.

Observaciones

* Consideramos que los resultados a nivel de métricas fueron parecidos, por lo tanto centramos el análisis en la mecánica de utilizar uno u otro approach.
* H2O provee una gran cantidad de documentación para iniciarse en su plataforma, ya sea a través de la interfaz Web (H2O Flow) o a través de las API en Python/R
* Utilizar la interfaz web es intuitivo y posee un amplio rango de opciones para manipular la información, ajustar hiperparametros y entrenar modelos.
* Es muy potente para realizar una exploración de los datos ya que entrena una gran cantidad de modelos y las condiciones de corte son varias. Por lo tanto podemos setear un máximo de tiempo (1hs) por ejemplo y la herramienta probara todos los modelos posibles en ese rango de tiempo.
* La información final es muy detallada y bien estructurada.
* Si comparamos el tiempo que nos lleva armar una notebook para probar modelos tradicionales o una CNN (con lo que significa elegir los hiperparametros o diseñar las capas de la red neuronal) contra hacer el mismo trabajo pero en la plataforma H2O, resulta notable cómo se pueden lograr muchas mas cosas en menor tiempo utilizando la plataforma de AutoML

Conclusiones

* H2O Flow provee variadas herramientas de fácil acceso para manipular el dataset de entrada y entrenar modelos. Ahora bien, esto no debe traducirse como que H2O es una herramienta que puede utilizar cualquier persona ajena al mundo de data science/machine learning.
* Consideramos que es una excelente herramienta para utilizar en un primer approach del modelo y luego en base a los resultados seguir iterando con el mismo H2O o frameworks tradicionales.