**Description:**

Trying to find the Hyperparameters for different dataset by running various Algorithms from H2OAutoML in python. We will be using Algorithms of types Common, Supervised, Unsupervised and Miscellaneous from H2O for different datasets from Kaggle/UCI. We will be running the Algorithms for different datasets on given a different point of time, trying to generate the Hyperparameters like rmse, mse, mean residual deviation, mae, and rmsle and store them for determining the best possible parameters to use. At the same time, we are going to save the model for which the data set has been run to avoid generating the model every time we run for the Kaggle dataset and instead, load it at the first time. The Hyperparameters that we generate for different periods (50,300,500,150sec) will be stored in a relational database. Also, we will be generating the JSON file for the respective dataset and model and store the same.

**Research:**

Understanding H20 package, hyperparameters and their calculations.

**Data Sources:**

Different datasets from Kaggle/UCI

**Algorithms:**

1. **Common**

* Quantiles
* Early Stopping

1. **Supervised**

* Cox Proportional Hazards (CoxPH)
* Deep Learning (Neural Networks)
* Distributed Random Forest (DRF)
* Generalized Linear Model (GLM)
* Gradient Boosting Machine (GBM)
* Naïve Bayes Classifier
* Stacked Ensembles
* XGBoost

1. **Aggregator**

* Generalized Low Rank Models (GLRM)
* Isolation Forest
* K-Means Clustering
* Principal Component Analysis (PCA)

1. **Miscellaneous**

* Word2vec

**References:**

<http://docs.h2o.ai/h2o/latest-stable/h2o-docs/index.html>

<https://www.kaggle.com/datasets>

https://archive.ics.uci.edu/ml/datasets.html

**Group:** Jaynee & Prabhu