PROJECT DESCRIPTION:

Goals and objectives: Hyperparameters are parameters that are specified prior to running machine learning algorithms that have a large effect on the predictive power of statistical models. Knowledge of the relative importance of a hyperparameter to an algorithm and its range of values is crucial to hyperparameter tuning and creating effective models. The hyperparameter database is created by running millions of hyperparameter values, over thousands of public datasets and calculating the individual conditional expectation of every hyperparameter on the quality of a model.

The hyperparameter database also uses these data to build models that can predict hyperparameters without search and for visualizing and teaching statistical concepts such as power and bias/variance tradeoff.

Project requirements: Basic understanding and working knowledge of machine learning algorithms.

Problems to be addressed: The hyperparameter database is a public resource with algorithms, tools, and data that allows users to visualize and understand how to choose hyperparameters that maximizes the predictive power of their models.

Potential pitfalls and challenges: Finding the right dataset.

BACKGROUND RESEARCH:

 H2O, Understanding hyper parameters and its optimization in data science.

ALGORITHMS AND CODE SOURCES:

Currently, the hyperparameter database analyzes the effect of hyperparameters on the following algorithms: Distributed Random Forest (DRF) Generalized Linear Model (GLM), Gradient Boosting Machine (GBM). Naïve Bayes Classifier, Stacked Ensembles, XGBoost and Deep Learning Models (Neural Networks).

DATA SOURCES:

- ✓ [List of datasets for machine learning research](#)
- ✓ [UC Irvine Machine Learning Repository](#)
- ✓ [Public Data Sets : Amazon Web Services](#)
- ✓ [freebase](#)
- ✓ [Google Public Data Explorer](#)
- ✓ [datahub](#)
- ✓ [data.gov](#)

REFERENCES:

- ✓ https://www.dataschool.io/15-hours-of-expert-machine-learning-videos/
- ✓ https://towardsdatascience.com/understanding-hyperparameters-and-its-optimisation-techniques-f0debba07568

PROJECT MEMBERS:

- ✓ Michelle Pradeep (NUID: 001422140)
- ✓ Seemanthini Jois (NUID: 001439942)
- ✓ Suraksha Jadhav (NUID: 001430368)