**DESCRIPTION OF DATASET**

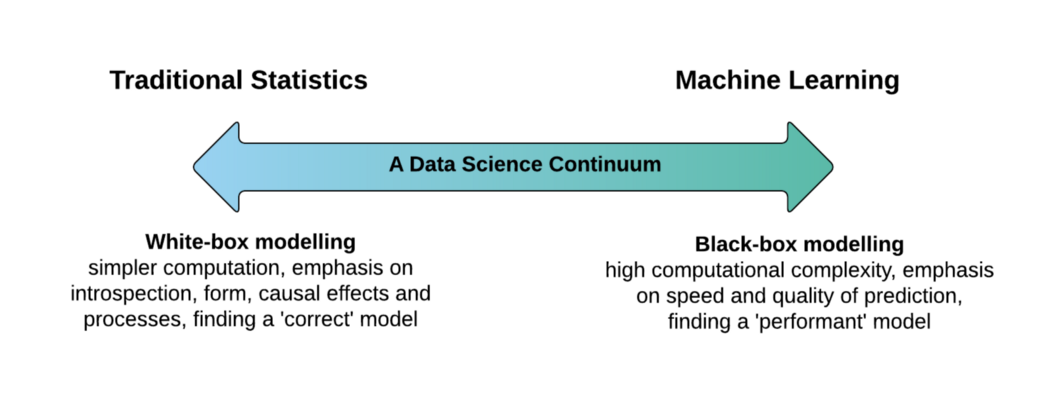
In [machine learning](https://en.wikipedia.org/wiki/Machine_learning), a common task is the study and construction of [algorithms](https://en.wikipedia.org/wiki/Algorithm) that can learn from and make predictions on [data](https://en.wikipedia.org/wiki/Data).

**THERE ARE THREE TYPES OF DATASETS**

* **Training Data**: is an extremely large dataset that is used to teach a [machine learning](https://www.techopedia.com/definition/8181/machine-learning-ml) model. Training data is used to teach [prediction models](https://www.techopedia.com/definition/14004/predictive-modeling) that use machine learning algorithms how to extract features that are relevant to specific business goals. For [supervised ML](https://www.techopedia.com/definition/30389/supervised-learning) models, the training data is labelled. The data used to train [unsupervised ML](https://www.techopedia.com/definition/30390/unsupervised-learning)models is not labelled.
* **Validation Data**: This Set provides an unbiased evaluation of a model fit on the training data set while tuning the model's [hyperparameters](https://en.wikipedia.org/wiki/Hyperparameter_(machine_learning)). It can be used for [regularization](https://en.wikipedia.org/wiki/Regularization_(mathematics)) by [early stopping](https://en.wikipedia.org/wiki/Early_stopping).
* **Test Data Set:** is a data set used to provide an unbiased evaluation of a final model fit on the training data set. If the data in the test data set has never been used in training (for example in [cross-validation](https://en.wikipedia.org/wiki/Cross-validation_(statistics))), the test data set is also called a holdout data set.

**DIFFERENCE BETWEEN WHITE BOX AND BLACK BOX DESIGN**

* **Accurate And ‘Black-Box’:**  
  Black-box models such as neural networks, gradient boosting models or complicated ensembles often provide great accuracy. The inner workings of these models are harder to understand and they do not provide an estimate of the importance of each feature on the model predictions, nor is it easy to understand how the different features interact.
* **Weaker And ‘White-Box’:**  
  Simpler models such as linear regression and decision trees on the other hand provide less predictive capacity and are not always capable of modelling the inherent complexity of the dataset (i.e., feature interactions). They are however significantly easier to explain and interpret.



**LIBRARY BASED NN DESIGN AND ITS RESULTS**

**IMPLEMENTED IN JUPYTER NOTEBOOK**

**WHITE BOX NN DESIGN AND ITS RESULTS**

**IMPLEMENTED IN JUPYTER NOTEBOOK**