**XG BOOST**

XG Boost is an algorithm that has recently been dominating applied machine learning and Kaggle competitions for structured or tabular data.

XG Boost is an implementation of gradient boosted decision trees designed for speed and performance.

**What is XG Boost?**

XG Boost is a software library that you can download and install on your machine, then access from a variety of interfaces. Specifically, XG Boost supports the following main interfaces:

* Command Line Interface (CLI).
* C++ (the language in which the library is written).
* Python interface as well as a model in scikit-learn.
* R interface as well as a model in the caret package.
* Julia.
* Java and JVM languages like Scala and platforms like Hadoop.

**XG Boost Features**

The library is laser focused on computational speed and model performance, as such there are few frills. Nevertheless, it does offer a number of advanced features.

### Model Features

The implementation of the model supports the features of the scikit-learn and R implementations, with new additions like regularization. Three main forms of gradient boosting are supported:

* **Gradient Boosting** algorithm also called gradient boosting machine including the learning rate.
* **Stochastic Gradient Boosting** with sub-sampling at the row, column and column per split levels.
* **Regularized Gradient Boosting** with both L1 and L2 regularization.

### System Features

The library provides a system for use in a range of computing environments, not least:

* **Parallelization** of tree construction using all of your CPU cores during training.
* **Distributed Computing** for training very large models using a cluster of machines.
* **Out-of-Core Computing** for very large datasets that don’t fit into memory.
* **Cache Optimization** of data structures and algorithm to make best use of hardware.

### Algorithm Features

The implementation of the algorithm was engineered for efficiency of compute time and memory resources. A design goal was to make the best use of available resources to train the model. Some key algorithm implementation features include:

* **Sparse Aware** implementation with automatic handling of missing data values.
* **Block Structure** to support the parallelization of tree construction.
* **Continued Training** so that you can further boost an already fitted model on new data.

XGBoost is free open source software available for use under the permissive Apache-2 license.

## Why Use XGBoost?

The two reasons to use XGBoost are also the two goals of the project:

1. Execution Speed.
2. Model Performance.

### 1. XGBoost Execution Speed

Generally, XGBoost is fast. Really fast when compared to other implementations of gradient boosting.

### 2. XGBoost Model Performance

XGBoost dominates structured or tabular datasets on classification and regression predictive modeling problems.

The evidence is that it is the go-to algorithm for competition winners on the Kaggle competitive data science platform.

## What Algorithm Does XGBoost Use?

The XGBoost library implements the [gradient boosting decision tree algorithm](https://en.wikipedia.org/wiki/Gradient_boosting).

This algorithm goes by lots of different names such as gradient boosting, multiple additive regression trees, stochastic gradient boosting or gradient boosting machines.

Boosting is an ensemble technique where new models are added to correct the errors made by existing models. Models are added sequentially until no further improvements can be made. A popular example is the [AdaBoost algorithm](http://machinelearningmastery.com/boosting-and-adaboost-for-machine-learning/) that weights data points that are hard to predict.

Gradient boosting is an approach where new models are created that predict the residuals or errors of prior models and then added together to make the final prediction. It is called gradient boosting because it uses a gradient descent algorithm to minimize the loss when adding new models.

This approach supports both regression and classification predictive modeling problems.