XGBoost and LightGBM

# General AdaBoost Algorithm

Diagram

Description automatically generated

Diagram

Description automatically generated

Shape

Description automatically generated

XGBoost 🡪 2 Individuals Research Paper

LightGBM 🡪 Microsoft Research

Data is less than **50k rows or 100k rows** then instead of ANNs we prefer GBDT (XGBoost, LightGBM)

# XGBoost

Gradient Boosting on a Steroid == XGBoost

Uses **CART** Decision Trees

Why XGBoost is one of the best?

1. **Parallelization**

Diagram

Description automatically generated with low confidence

1. **Hardware Support:**
   1. GPU (Utilize All cores – 1)
   2. CPU (Utilize All cores – 1)
2. **Parameter Tuning:**
   1. Inbuilt CV and Tuning Approaches
3. **Missing Values and Outliers:**
   1. Automatic Treatment
4. **Depth First and Leaf First Model**

Diagram

Description automatically generated

# LightGBM

1. Default use is **GOSS** decision Tress
2. GOSS trees have drop-outs options similar to ANN.
3. Only has option of Leaf First Approach
4. Not recommended for data less that 10k
5. Other points similar to XGBoost

**Few rules:**

* Less than 10k data, use XGBoost
* Less than 100K data, use XGBoost or LightGBM
* More than 100K data, try ANN, XGBoost or LightGBM
* Billions or rows 🡪 ANNs