What is XGBOOST?

XGBoost is one the popular libraries known for its well-optimized code

For example :

- if you have numerical features f_i
    - instead of trying all the values for thresholding,
    - It builds a histogram of data and uses simple rules like quartiles and percentiles to make thresholding.
- It also does multi-core optimization (parallelization)
    - it'll compute each branch of a base learner on a different core to speed up the process.

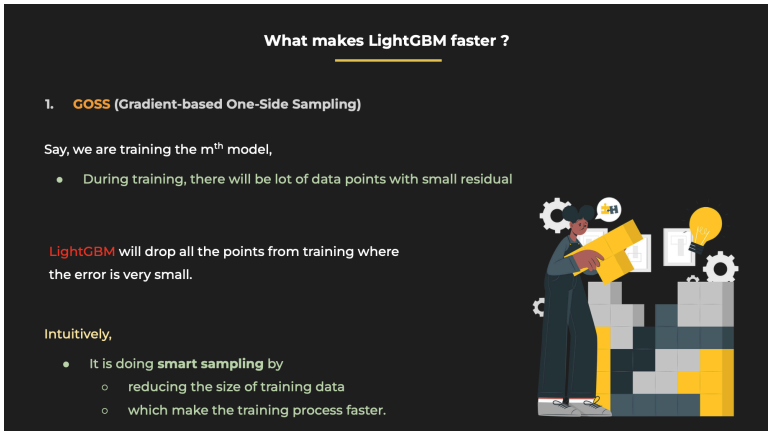
Some commonly used hyper params of XGBOOST

- Number of estimators (M)
- Depth
- v: learning rate
- Col sampling/ row sampling

What is LightGBM?

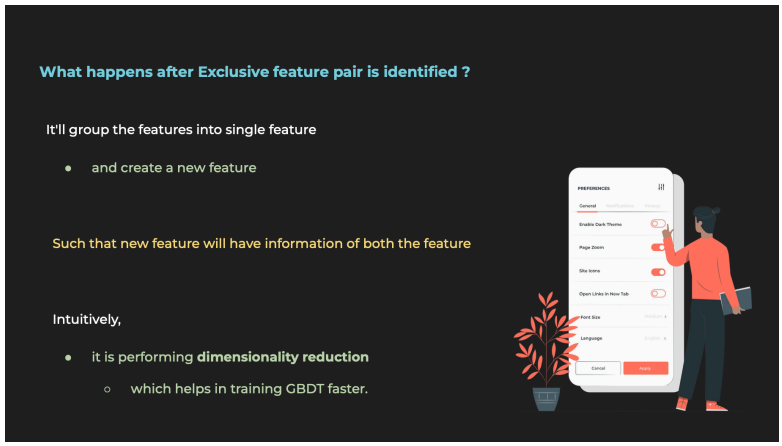
It has major 2 optimizations over XGBOOST

1. GOSS(Gradient-based one-side sampling) operates by selectively sampling instances based on the gradient information obtained during training



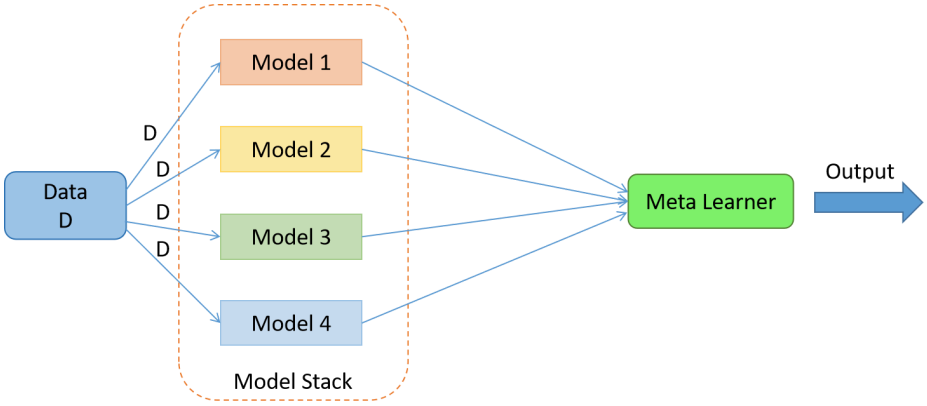
- In addition to preserving instances with large gradients, GOSS also applies one-side sampling to reduce the number of instances with small gradients.
    - It randomly samples a fraction of instances with small gradients,

2. Exclusive Feature Bundling (EFB)
    - tries finding feature pairs s.t they are exclusive



What is Stacking?

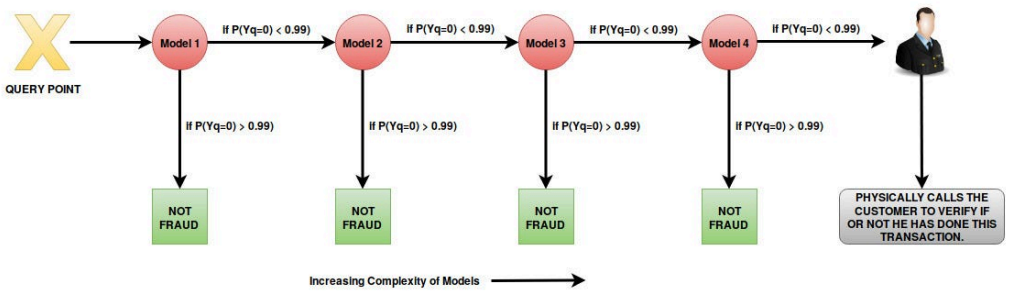
Here, we are taking the outputs of the perfectly built models and stacking them together to train a Meta-classifier to get the final output



What is Cascading?

- Cascading is a type of ensemble model where multiple base learners are organized into a sequence such that they make a decision, and the new

models are trained only on the datapoints that were wrongly classified by the previous base model
- Very useful when Sensitivity is crucial



DIFFERENT STAGES OF QUERING CASCADE CLASSIFIERS IN A FOUR MODEL CASCADE SYSTEM

Why is GBDT used more often than RF?

1. Any differentiable loss function can be used
2. GBDT has a cheaper run time because
    - the base learners are shallow and
    - The random forest has deeper trees and
    - the number of trees to train in GBDT is comparatively less

When should we use Cascading and stacking?

Cascading is used when the risk or cost of mistakes is high, and the data is highly imbalanced.

- Like fraud transaction detection in Amazon

What about the explainability of the model?

- We make sure that every model is explainable so that we can explain the output using these models
- We will see a few algorithms, like **LIME and SHAP** which can explain any black box algorithm after a few lectures in Deep Learning.

Stacking is mostly seen in Kaggle competitions, not so much in the real world.