

For instance, the average impurity loss of feature  $j$  on 10 subtrees of GBDT is calculated by formula 1.

$$\hat{L}_j^2 = \frac{1}{M} \sum_{m=1}^M \hat{L}_j^2(T_m) \quad (1)$$

Where,  $M$  is the number of subtrees and particularly,  $M = 10$  in our experiment.  $\hat{L}_j^2(T_m)$  is the impurity loss of feature  $j$  on subtree  $m$  that can be figured as follow.

$$\hat{L}_j^2(T_m) = \sum_{t=1}^{N-1} \hat{i}_t^2 I(v_t = j) \quad (2)$$

Where,  $N$  is the number of nodes of tree and  $N - 1$  is the number of non-leaf nodes.  $v_t$  represents features related to node  $t$ .  $\hat{i}_t^2$  is the square of reduced impurity after splitting node  $t$ .