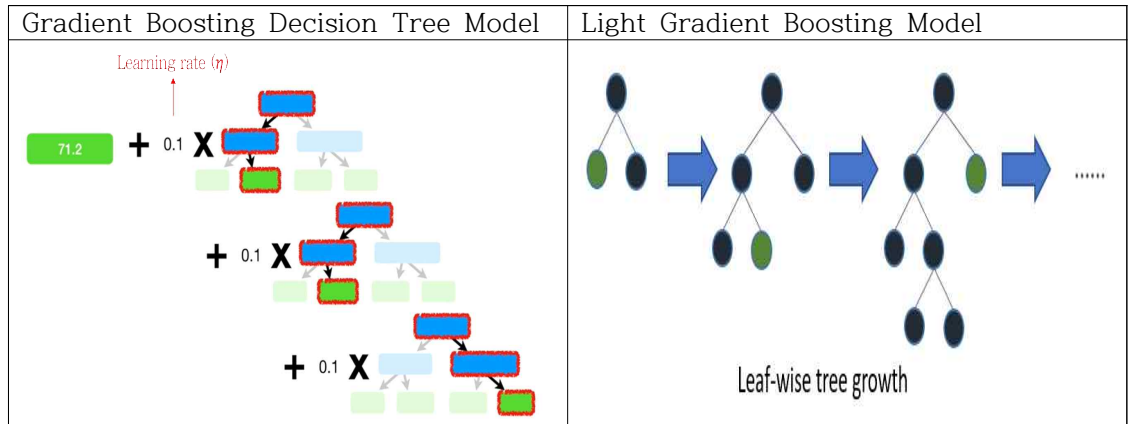


Gradient Boosting Decision Tree Model & Light Gradient Boosting Model

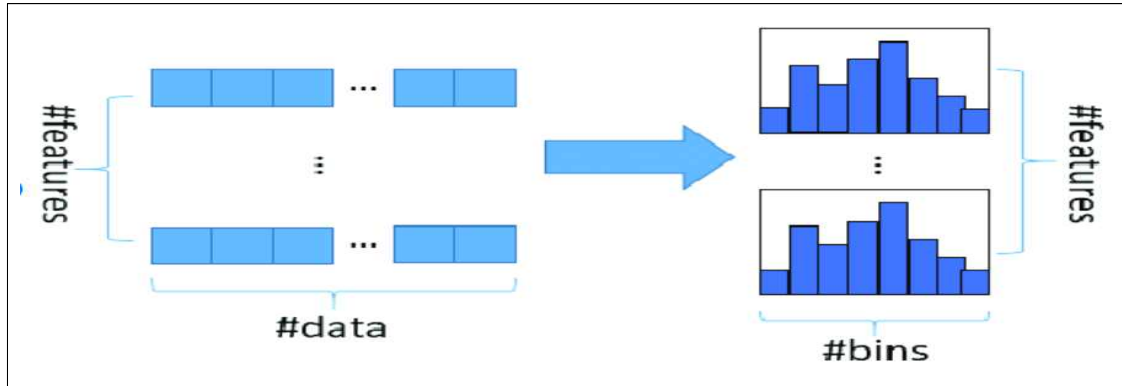


[표 1] GBT 모델과 LGB 모델의 작동원리

Gradient Boosting(GBT) 모델은 트리를 생성하여 각 output의 평균이 leaf가 가지는 값이 되며, 이후 잔차를 계산한다. 이 과정을 반복하여 이전 트리의 잔차를 점차 줄여나가는 기법이다. GBT 모델은 트리의 깊이를 뜻하는 max\_depth의 값이 3~6인 얇은 의사결정 나무일 때 높은 예측력을 보이는 특징이 있다[1].

GBT 모델의 경우, 트리 분할 방식이 균형 트리 분할 (level-wise) 방식을 채택하기 때문에 모델의 parameter를 조절할 때 과적합 문제에 대하여 더 강한 구조를 가지는 장점이 있다. 그러나, 균형 트리 분할 방식이기 때문에, 훈련 시간이 오래 걸리며, 메모리 소비가 크다는 단점이 존재한다[2]. 이를 극복하여 만들어진 알고리즘이 LightGBM이다. LightGBM은 GBT 모델의 가장 큰 단점인 데이터 처리시간을 GOSS(Gradient Based One Side Sampling)와 EFB(Exclusive Feature Bundling) 기법으로 보완한 알고리즘으로[3], 리프 중심 트리 분할 (Leaf-wise) 방식을 통해, max delta loss(최대 손실 값)를 가지는 리프 노드를 지속적으로 분할하기 때문에 비대칭 트리가 만들어지게 된다. 이러한 LightGBM 모델은 빠른 학습 및 예측 수행 시간, 더 적은 메모리 사용량, 더 좋은 정확성, 병렬 계산 및 분할, GPU 학습지원, 대규모 데이터 처리 유용 등의 특징을 가지고 있다. 단, LightGBM을 사용하기 위해서는 데이터의 개수가 10,000개 이상이어야 하는데, 만약 데이터의 수가 적을 경우, 과적합의 문제가 발생할 가능성이 높다.

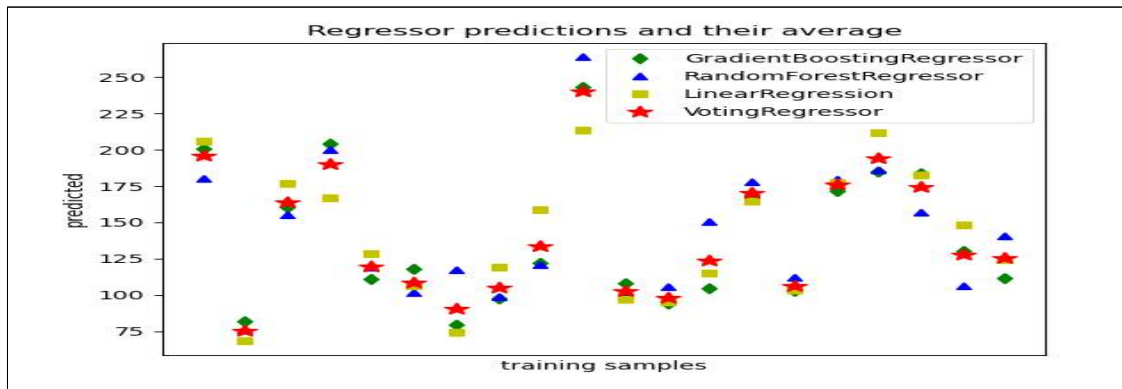
## Histogram-Based Gradient Boosting Model



[표 2] Histogram based Gradient Boosting 모델의 작동 원리

Histogram-Based Gradient Boosting(HGBT) 모델은 LightGBM과 매우 유사한 모델로, LightGBM의 특징과 더불어 결측치에 대한 imputer가 필요하지 않다는 추가적인 특징을 가지고 있다. 다만, LightGBM과 달리 전처리 과정에서 사용하는 binning을 의사 결정 트리 알고리즘에도 적용하여 알고리즘의 속도를 높인다는 것에 차이가 있다. 이때, 모델 스스로 최적의 bins를 찾아내 고려할 분할점의 수를 줄이며(일반적으로 최대 bins는 256), 정렬된 연속 값에 의존하지 않은 데이터 구조를 활용할 수 있다. HGBT 모델의 parameter 중 max\_iter는 GBT 모델의 parameter 중 n\_estimator와 같은 역할을 한다. 예측을 위한 훈련 데이터의 양이 많을 경우, 데이터의 크기를 줄이기 위한 일반적인 방법은 다운 샘플링 기법을 사용한다 [2]. 하지만 다운 샘플링을 할 경우, 데이터를 손실한다는 단점이 존재하는데, HGBT 모델의 경우, 모델 학습의 속도가 빠르므로, 다운 샘플링이 필요 없고, 그 결과 데이터 손실을 최소화한다는 큰 장점이 있다.

## Voting Regressor



[표 3] 여러 모델들의 예측 값

Voting 기법의 경우, 한 모델에서 다른 sample 조합을 사용하는 bagging 기법과 달리, 서로 다른 모델 간의 조합을 통해 평균 예측값을 반환하는 기법이다. 이 기법의 경우 한 모델에 대한 약점을 극복할 수 있는 장점이 있다.

<https://tyami.github.io/machine%20learning/ensemble-4-boosting-gradient-boosting-regression/> (GBT 모델 사진)

<https://nurilee.com/2020/04/03/lightgbm-definition-parameter-tuning/> (LGBM 사진)

<https://www.analyticsvidhya.com/blog/2022/01/histogram-boosting-gradient-classifier/> (histogram-based GBT 사진)

#### 참고 문헌

[1]B. Ilyasov, E. Makarova, V. Martynov, E. Zakieva, E. Gabdullina and M. Yusupov, "Application of Gradient Boosting Algorithm for Predicting Equipment Failures," 2022 VI International Conference on Information Technologies in Engineering Education (Inforino), 2022, pp. 1-5, doi: 10.1109/Inforino53888.2022.9783011.

[2]Ke, Guolin, et al. "Lightgbm: A highly efficient gradient boosting decision tree." *Advances in neural information processing systems* 30 (2017).

[3]장승일, 광근창.(2019).XGBoost와 LightGBM을 이용한 안전 운전자 예측 성능 비교.한국정보기술학회 종합학술발표논문집,( ),360-362.