RandomForest 파라미터 변경 시 성능 변화

효율적인 모델 튜닝을 위한 파라미터 영향 분석

n_estimators
트리 개수

max_depth
트리 깊이

min_samples_split
노드 분할 최소 샘플

min_samples_leaf
리프 최소 샘플

n_estimators (트리 개수)

역할

랜덤포레스트가 사용하는 결정 트리의 개수

↑ 값 증가 (예: 100 → 300)

- 정확도 ↑
- 예측 안정성 ↑
- 학습 시간 증가

↓ 값 감소 (예: 300 → 50)

- 학습 빠름
- 정확도 ↓ 가능
- 모델 변동성 ↑

n_estimators 변경 시 모델 성능 변화 추세

max_depth (트리 깊이)

역할

트리가 끝까지 자랄 수 있는 최대 깊이

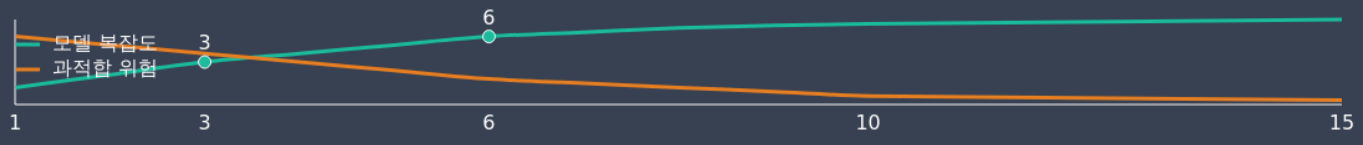
↑ 값 증가 (6 → 15)

- 모델 복잡도 ↑
- 훈련 정확도 ↑
- 과적합 위험 증가

↓ 값 감소 (6 → 3)

- 모델 단순화
- 학습 부족(언더피팅) 가능
- 훈련 정확도 ↓

max_depth 변경 시 모델 복잡도 및 과적합 위험



min_samples_split

역할

하나의 노드를 분할하기 위한 최소 샘플 수

↑ 값 증가 (2 → 5 → 10)

- 노드가 잘 안 나눔
- 모델 단순화
- 과적합 방지

↓ 값 감소 (5 → 2)

- 더 세밀하게 나눔
- 모델 복잡도 ↑
- 과적합 위험 ↑

min_samples_split 변경 시 트리 분할 영향

min_samples_leaf

역할

리프(leaf)에 최소 몇 개의 샘플이 있어야 하는지

↑ 값 증가 (1 → 5 → 10)

- 작은 샘플로 판단 X
- 과적합 ↓
- 너무 크면 성능 ↓ 가능

↓ 값 감소 (5 → 1)

- 리프를 더 세밀하게 만듦
- 복잡도 ↑
- 과적합 ↑

min_samples_leaf 변경 시 모델 성능 변화 추세

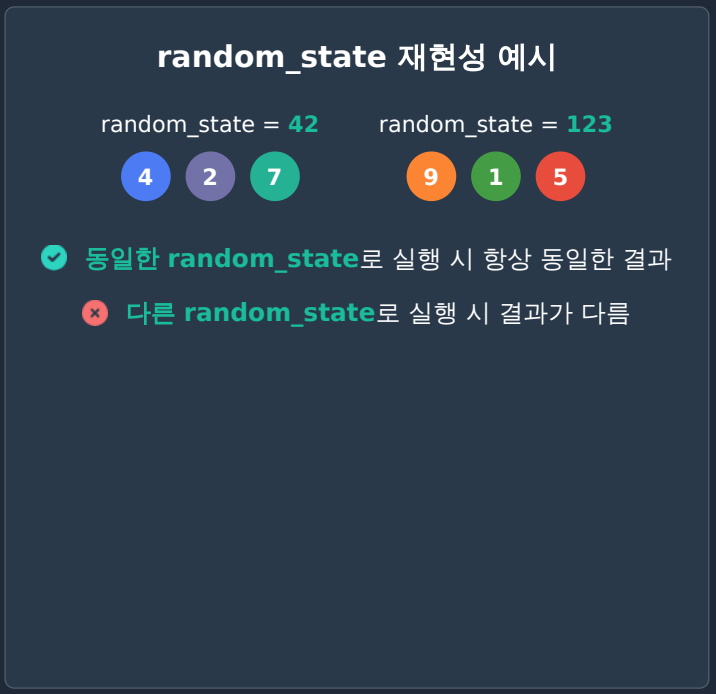
1 적은 샘플 → 5 중간 샘플 → 10 많은 샘플

random_state

역할
모델의 랜덤 요소를 고정

결과 재현성을 위한 설정
동일한 데이터와 파라미터로 항상 동일한 결과를 보장

성능에 직접 영향 없음
모델의 예측력에는 영향을 주지 않음



튜닝 요약 표

RandomForest 파라미터 변경 시 성능과 속도, 과적합/언더피팅 영향을 정리한 표입니다.

파라미터	값을 키우면	값을 줄이면
n_estimators	↑ 성능↑ , 속도↓	↓ 성능↓ , 속도↑
max_depth	↑ 과적합↑	↓ 언더피팅↑
min_samples_split	↓ 과적합↓	↑ 과적합↑
min_samples_leaf	↓ 과적합↓	↑ 과적합↑
random_state	= 동일 결과 유지	= 동일 결과 유지

↑증가 ↓감소 =변화 없음

상황별 튜닝 전략

모델 성능이 의도한 대로 나오지 않을 경우, 적절한 파라미터 조정을 통해 개선할 수 있습니다

모델이 과적합일 때

- **max_depth ↓**
  트리 깊이 제한을 낮춰 모델 복잡도 감소
- **min_samples_leaf ↑**
  리프에 필요한 최소 샘플 수 증가
- **min_samples_split ↑**
  노드 분할에 필요한 최소 샘플 수 증가

모델이 언더피팅일 때

- **max_depth ↑**
  트리 깊이를 늘려 모델 복잡도 증가
- **n_estimators ↑**
  트리 개수를 늘려 모델 안정성 향상
- **min_samples_leaf ↓**
  리프에 필요한 최소 샘플 수 감소

팁: 과적합과 언더피팅을 구분하는 가장 좋은 방법은 검증 세트 성능을 모니터링하며, 학습 곡선(learning curve)을 분석하는 것입니다.