

모델별 성능 시각화 및 예측 결과 인사이트

1. 모델별 성능 시각화 인사이트

- **XGBoost(튜닝), LightGBM(튜닝) 모델의 RMSE가 각각 37.47, 38.40으로 가장 낮았습니다.**
이는 자전거 대여량 예측에서 평균 오차가 최소였음을 의미합니다.
곧, 실제 대여량과의 차이가 가장 작아 가장 현실적인 예측이 가능하다는 것을 보여줍니다.
- **RandomForest, GradientBoosting, LightGBM 등 기본(default) 모델은 RMSE가 41 이상으로 더 높게 나타났습니다.**
하이퍼파라미터 튜닝 전과 후의 오차 차이가 명확하게 드러났으며,
특히 트리 기반 부스팅 모델은 튜닝을 통해 실질적인 성능 개선 효과가 있음을 알 수 있습니다.
- **R² Score(결정계수) 지표를 보면, GradientBoosting(튜닝), XGBoost(튜닝), LightGBM(튜닝) 모델이 모두 0.955~0.958 수준으로, 데이터의 변동성 중 95% 이상을 설명할 수 있음을 확인했습니다.**
- **전반적으로 튜닝된 부스팅 계열(XGBoost, LightGBM, GradientBoosting) 모델의 성능이 가장 우수하게 나타났으며,**
기본 모델 대비 튜닝을 통해 예측력(오차 감소, 설명력 증가) 모두 실제로 개선되는 점이 명확히 시각적으로 드러납니다.

2. 예측 결과(count) 해석 및 인사이트

- **최종 count 예측 결과(submission.csv)는 튜닝된 XGBoost, LightGBM 등 최고 성능 모델의 예측값을 바탕으로 생성되었습니다.**
예측값은 음수 보정(0 미만은 0으로 처리) 등 실무에서 발생 가능한 예외 상황까지 고려하여 제출 파일로 작성하였습니다.
- **주요 인사이트로는 '시간대', '온도', '요일', '휴일여부' 등이 예측에 큰 영향을 미치는 것으로 나타났으며,**
실제 운영에서는 이 변수들을 중심으로 수요예측, 자전거 배치, 마케팅 타이밍 등을 최적화할 수 있습니다.

3. 결론 및 실무 제안

- 실제 서비스 적용 또는 경진대회 참여 시, 트리 기반 부스팅 계열(XGBoost, LightGBM 등) 모델과 하이퍼파라미터 튜닝은 반드시 고려해야 할 핵심 전략임이 실험적으로 검증되었습니다.
- 예측력(오차 감소)과 설명력(결정계수 향상)을 모두 높일 수 있는, 실무 최적 솔루션임을 확인할 수 있었습니다.
- 변수 중요도(Feature Importance) 해석을 병행하여, 수요 예측 및 정책/운영 의사결정에도 적극 활용할 수 있습니다.
- 실무 현장에서는, 모델 선정·튜닝뿐 아니라 실제 데이터 흐름(파생변수 생성, 이상치/음수 보정, 결과물 활용 방식)까지 전 주기에 걸친 관리가 중요합니다.

핵심 요약

- 트리 계열 부스팅 모델과 튜닝은 자전거 대여량 예측 문제에서 실제로도, 데이터 적으로도 가장 높은 신뢰성과 성능을 보여줍니다.
- 시각화와 성능 비교, 예측값의 실무적 검증은 효과적인 의사결정, 운영 효율화, 정책 수립의 강력한 근거가 됩니다.

