

10월 22일 공유

1. 모델링 결과

- 파라미터 튜닝까지 진행
 - 파라미터 튜닝은 AutoML 패키지 pycarrot을 통해 여러가지 파라미터 조합 테스트와 개인적인 경험을 바탕으로 진행.
- 결과 - MAE 성능 순:
 - **LinearRegression > LightGBM > RandomForest > DecisionTree ≥ ExplainableBoostingModel**

2. 결과 해석

- LinearRegression 모델이 가장 좋은 성능을 보인다.
 - 가장 투박하면서 심플한 LinearModel의 성능이 좋은 이유는?
 - 학습 데이터가 모두 연속형 변수이다.
 - ⇒ 우리가 보유한 데이터가 뉴스 건수, 선호도 등 모두 연속형 변수이다.
 - ↔ {성별 : '남', '여'} 와 같은 이산형 변수가 존재하지 않는다.
- ExplainableBoostingModel의 경우 가장 안좋은 성능을 보이는데,,,, 현재 데이터셋에 는 맞지 않는 느낌

3. 결론

최종 모델로는 **LinearRegression** or **LightGBM** 모델을 택하는 것이 좋아보입니다!

모델링 결과_final.xlsx

