**수리비 예측 모델 제작 보고 2**

박성완

1. **새로운 Feature 추가**

* **HQ :** 수리 시간 (단, 도장 수리의 경우 무조건 100으로 되어 있음, 다른 수리들은 평균적으로 0~10 사이로 되어 있는듯)
* **REPAIRDAY :** 수리를 맡긴 날
  + **integer feature** : MILEAGE, FIRSTDAY, REPAIRDAY, HQ, PRICE
  + **categorical feature** : PART, SEVERITY

1. **추가 데이터셋 생성**

* 저번주는 전체 데이터셋을 사용하지 않았지만 이번에는 83만개의 데이터 생성

텍스트이(가) 표시된 사진

자동 생성된 설명

1. **이상 데이터 수정**

* Part에 “신품가”가 들어갈 경우 수리비가 무조건 0이 나오는 것을 확인,

텍스트이(가) 표시된 사진

자동 생성된 설명

* Part에 있는 가격을 빼고 Price column으로 이동.

1. **Categorical feature 압축**

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

1. **LabelEncoding 통한 categorical feature 처리**

저번주는 PART, SEVERITY feature feature에 one-hot-encoding을 진행하여 feature의 개수가 너무 많아지는 문제가 발생, 이번주는 LabelEncoding을 진행하여 최종 feature의 수를 6개로 줄임.

1. **MinmaxScaler를 통한 정규화 실행**

텍스트이(가) 표시된 사진

자동 생성된 설명

저번주는 categorical feature도 같이 scaling 했지만 이번에는 integer feature들로만 Scaling 진행.

1. **GradientBoosting, RandomForestRegressor Model 학습**

텍스트이(가) 표시된 사진

자동 생성된 설명

매우 정확한 결과가 나오는 것을 볼 수 있음.

1. **실제 데이터로 테스트**

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

다양한 가격대의 차를 넣어 봤는데도 성능이 괜찮은 것을 확인 할 수 있음.

1. **차후 개선 방안**

* 현재는 Train,Test 데이터를 랜덤 샘플링하고 있지만, 더 좋은 Sampling 방식을 찾아서 진행해볼 예정.
* 현재는 사용하지 않는 COMPANY, CARNAME, MODELTYPE으로 소형/중형/대형차를 나누고 외제차,국산차를 나누는 방법으로 새로운 feature를 적용해볼 예정.