0 안녕하세요 electric shock팀 발표를 맡은 안은지입니다. 저희 팀은 문영재, 서강석, 조은겸, 저를 포함해 4명이고, 팀명에서 눈치채신 분도 있으시겠지만, 주제는 전력사용량 예측입니다. 저희는 ‘2023년 전력사용량 예측 AI 경진대회’ 공모전과 프로젝트를 동시에 진행하였습니다. 공모전의 과제는 8월의 마지막주 예상 전기사용량으로 미래 예측 데이터가 필요합니다.

1 본격적인 데이터 탐색에 앞서 선행연구와 공모전 사례들을 간단히 찾아봤습니다. 그중 딥러닝 방식인 LSTM이 가장 뛰어난 것으로 나타났고, 공모전에서는 부스팅 모델들을 많이 채택한 것을 알 수 있었습니다.

2 또 기상청이 발표한 자료에 여름에는 최저기온, 운량, 체감온도와 전력사용량의 상관성이 높은 것으로 나타났습니다. 물론 여기는 기상 변수 외에는 고려하지 않았으나, 주최측에서 준 테스트 데이터 또한 기상요소만 있었기에 변수는 기상요소를 우선적으로 고려하였습니다. 하지만 전체 건물로 봤을 때 요소간의 상관관계가 예상 외로 강하지 않았습니다.

3 우선 그래프로 데이터를 파악해보기로 했습니다. 지금 보시는 그래프와 같이 일시, 건물 유형별로 전력사용량을 비교했습니다. 차이가 제법 뚜렷하게 보이는 걸 알 수 있습니다. 같은 건물 유형이면 전력사용량이 비슷하지 않을까라는 생각에 같은 유형 건물도 비교했습니다. 놀랍게도 동일한 유형에서도 차이가 난다는 사실을 발견했습니다.

4

소감

첫 팀플인데 정말 좋은 조원들을 만나 혼자 하는 것보다 많이 공부하고, 편안하게 프로젝트를 진행할 수 있어 무척 좋았습니다. LSTM을 너무나 하고 싶었지만 R에서 할 수가 없어 몹시 슬펐고, LSTM이 아니더라도 keras, tensorflow 등 다른 라이브러리도 구동에 어려움을 겪으면서 그만큼 얼른 파이썬을 배우고 싶다는 생각도 강해졌습니다. 또 프로젝트를 진행하면서 가시적인 도메인 지식과 파생변수 생성의 중요성을 깨달았습니다. 부스트 모델들이 파라미터가 정말 많은데 주요 하이퍼파라미터를 약간씩 조정해봤지만 큰 차이를 내지는 못하였습니다. 후에 파이썬으로 할 때는 optuna로 최적의 파라미터 값을 찾고 싶습니다.