기말프로젝트 발표를 하게 된 김효재, 이원희 조의 ㅁㅁㅁ 입니다. 발표 시작하겠습니다.

차례입니다. 발표는 주제에 대해 간단하게 설명하고, 데이터 클린징, 훈련 및 평가순서로 진행됩니다.

저희 조는 날씨정보를 가지고 모기의 개체수를 예측해보는 것이 목표입니다. 모기의 개체수 정보는 서울시 디지털 모기 측정기를 통해 측정된 모기 수 자료에서 가져왔고, 날씨정보는 서울시의 일일 강수량, 습도, 기온 자료를 기상 자료 개방 포털에서 가져왔습니다.

다음은 데이터 클린징입니다.

DMS 측정치에 결측치가 있는 경우가 있었고 이유를 찾아보니 기기결함이나 주변의 정전 같은 이유로 측정이 멈추는 경우가 있었습니다. 이런 경우 복구하는데 2~3일정도가 소요되며 해당 기기간동안 결측치가 발생했습니다. 저희는 이런 결측치가 있는 부분을 삭제하고 진행했습니다.

비는 매일 내리지 않기 때문에 강수량이 없는 일자에는 강수량 칸이 비어 있었습니다. 이 부분은 비가 안왔다는 뜻이므로 해당일자 강수량 정보에 0을 입력했습니다.

다음은 훈련 및 평가입니다.

저희는 사이킷 런에 있는 선형회귀, SVM 회귀, 다층신경망 회귀를 이용해서 훈련을 진행했습니다.

다만, 훈련을 진행하기 전에 자료들의 크기나 단위가 다르기 때문에 이 자료들을 피처 스케일링을 통해서 피쳐의 크기를 비슷한 수준으로 조정하고 훈련을 진행했습니다.

훈련 코드와 결과 사진을 같이 보시겠습니다.

먼저 선형회귀입니다. RMSE가 약 1696이고, R2스코어값은 0.325가 나왔습니다.

SVM 회귀입니다. RMSE는 약 1826이며, R2스코어 값은 0.217임을 확인할 수 있었습니다.

마지막으로 MLP회귀입니다. RMSE는 약 1728이며, R2 스코어 값은 0.299 였습니다.

훈련을 진행하고 나서 보면 선형회귀의 결과가 가장 우수했다는 것을 알 수 있었습니다.

이번 발표를 준비하면서 아쉬웠던 것은, 저희가 초반 동영상 녹화를 하며 참고했던 것 중 위치와 관련된 피쳐가 있다면 좀 더 정확한 값이 나올 것이다 라는 내용이 있었지만 강이나 습지 근처의 모기 포집량을 정확하게 알 수 없어 위치와 관련된 피쳐를 찾아보지 못했다는 점 입니다.

관심이 있으시다면 관련된 논문을 발표자료의 참고란에 추가해 놓았으니 참고하시길 바랍니다. 감사합니다.