

「Do Learn Do Run」 프로그램 연구활동 주간보고서

연구분야	데이터 시각화	지도교수	장 윤 (확인)
멘티명	학과(전공) <u>지능기전공학부 스마트기기전공</u> 학번 <u>19011755</u> 성명 <u>이지민</u>		
참여연구과제명	한강 공원 주차장 이용자수 예측 시스템 설계		
활동기간	2021.11.29.월~2021.12.3.금		

주요 연구 활동 내용

2021.11.30.화

직접 인공지능 모델을 구축하고 학습시킨 데이터를 시각화해보는 과정을 배우기 위해 'covid-19 시각화 시스템 설계' 주제에서 '한강 공원 주차장 이용자수 예측 시스템 설계' 주제로 변경함.

2021.12.1.수 ~ 2021.12.3.금

서울 열린데이터 광장의 한강공원 주차장 일별 이용 현황 데이터

(<https://data.seoul.go.kr/dataList/OA-21084/S/1/datasetView.do>) 와

기상청 기상자료개방포털의 기온분석 데이터 (<https://data.kma.go.kr/stcs/grnd/grndTaList.do?pgmNo=70>) 를 융합하여 데이터를 가공하였음. 한강공원 주차장 월별 이용 현황 데이터를 보면 11월 이용자수는 6만여명, 6월 이용자수는 11만여명으로 거의 2배 차이가 발생. 이를 통해 이용자수는 기온에 따른 영향이 큰 것으로 해석함.

이는 코로나19로 인해 많은 사람들이 모이는 장소를 기피하는 요즘, 조금이라도 사람이 적을 때 한강공원을 이용할 수 있다는 장점이 있음. 한강공원 이용객이 아닌, 한강공원 주차장 이용자수를 예측한 이유는 서울시 한강공원 이용객 현황 통계자료도 존재하지만 이는 년 단위의 데이터이기 때문에, 일 단위로 예측할 수 없어 부적합하다고 판단하였기 때문임.

이러한 데이터를 이용해 한강공원 주차장 이용자수를 예측할 수 있는 인공지능 모델을 구축함.

train과 test의 비율을 30%로 하여 약 6천개의 데이터를 train 데이터로 사용함. 데이터에 문자열도 존재하여 LabelEncoder를 사용하여 전처리 해주었고, 데이터의 각 단위가 달라 정규화 시켜주기 위해 StandardScaler를 사용해주었음. 모델은 linear 레이어를 4개 쌓아주었으며, 각 feature는 1024로 맞춰줌. xavier_normal로 초기화 시켜주었으며, 0.2의 비율로 dropout 시켜줌. learning rate는 0.001로 총 3000번 반복하여 학습시켜줌. 이 결과 r2_score로 정확도를 측정해보았을 때, train은 0.9695로 test는 0.96215로 성능이 나왔음.

모델학습 시켜 얻어낸 데이터를 이용해 추후 데이터 시각화를 진행할 계획임.