|  |  |
| --- | --- |
| 교육 제목 | AI 프로젝트 기반 S/W 전문가 양성과정 |
| 교육 일시 | 21.11.11 |
| 교육 장소 | 자택 |
| **교육 내용** | |
| 오전 | 1. Machine Learning    1. ‘살짝’ 맛보기    2. https://archive.ics.uci.edu/ml/index.php    3. drop\_first = yes, no column 두개 생기는데 처음 column을 drop 처음 column은 필요가 없음, 하나의 컬럼으로 필요한 정보를 모두 확인할 수 있음 2. PANDAS DATAFRAME 데이터 전처리 과정    1. 시리즈 함수 매핑 : Series.apply(함수명)    2. 데이터프레임 함수 매핑 : DataFrame.applymap(함수명)    3. 데이터프레임 각 열에 함수 매핑 : DataFrame.applymap(함수명, axis = 0)    4. 데이터프레임 각 행에 함수 매핑 : DataFrame.applymap(함수명, axis = 1)    5. 데이터프레임 함수 매핑 : DataFrame.pipe(함수명)  -> 반환하는 종류에 따라 반환 타입이 정해짐 3. DataFrame 열 재구성    1. column 리스트를 만들어서 정렬하고 다시 부여해줌 col\_list = list(df.columns.values) df = df[sorted(col\_list)]    2. col\_name = [“a”, “b”, “c”, “d”]    3. df = df[col\_name] 4. boolean indexing    1. mask = 조건식 / 많이 씀    2. Series.isin(찾고자 하는 값 / []) 5. DF 합치기 concat    1. pd.concat() 축을 기준으로 있으면 넣고 없으면 만들고 빈자리는 null    2. concat([df1,df2]): 기존의 인덱스를 그대로 가져옴, 행으로 붙이기, 없는 컬럼값은 null, parameter ignore\_index = True : 인덱스 새로 부여    3. concat([df1,df2], axis = 1)    4. join 있다. DBMS 개념 6. DF 합치기 merge    1. pd.merge() 정렬 후 가져옴    2. pd.merge() 같은 column이 있으면 해당 column의 값이 같은 자료를 합침 / default how = “inner” 7. DF 합치기 join    1. df1.join(df2, how = default “left”) df1을 기준으로 df2와 합치는데 df2에 없으면 null처리 |
| 오후 | 1. DF 합치기 groupby()    1. 분할(split) -> 적용(apply) -> 결합(combine)    2. DF.groupby(“컬럼명”) = 컬럼명 그룹을 DF로 만들어 줌    3. 변수 = DF.groupby(“컬럼명”) 변수.집계함수() = 컬럼명을 인덱스로 갖고 집계함수 값을 갖는 DF가 나옴    4. DF.groupby(“컬럼명1”).get\_group(“컬럼명2”) -> 컬럼명1으로 group을 나누고 컬럼명1을 인덱스로 갖는 DF가 나오고 그 group의 컬럼명2의 DF를 가져옴    5. DF. agg(함수 or [함수리스트]) = df에 함수 | 함수 리스트 적용    6. DF.agg({컬럼1:함수명, 컬럼2:함수명}) DF.agg({“col1”: [“min”, “max”], “col2”: “std”, “col3”: “mean”}) col1에 최소, 최대값, col2 표준편차, col3 평균    7. 실습    8. DFgroupby.컬럼.transform(함수명)    9. 실습    10. DF.xs(“컬럼”, level = “class”) 멀티 인덱스일 때 subindex에 접근 2. pivot 피벗    1. 엑셀의 피벗테이블 3. 과정    1. 전처리    2. test 3 : train 7 분리    3. 모델 train test    4. test 모델로 검증 |